FUNDAÇÃO OSWALDO CRUZ

INSTITUTO GONÇALO MONIZ

FIOCRUZ

Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa

TESE DE DOUTORADO

CARACTERIZAÇÃO MOLECULAR DA ORF-I DO HTLV-1 PROVENIENTE DE PACIENTES COM DIFERENTES PERFIS CLÍNICOS.

FERNANDA KHOURI BARRETO

Salvador – Bahia
2017

FUNDAÇÃO OSWALDO CRUZ

INSTITUTO GONÇALO MONIZ

Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa

CARACTERIZAÇÃO MOLECULAR DA ORF-I DO HTLV-1 PROVENIENTE DE PACIENTES COM DIFERENTES PERFIS CLÍNICOS.

FERNANDA KHOURI BARRETO

Orientador: Dr. Luiz Carlos Júnior Alcântara

Co-orientadora: Dr[a] Maria Fernanda de Castro Amarante

Tese apresentada ao Curso de Pós-Graduação para obtenção do grau de Doutor.

Salvador – Bahia
2017

"ORF-1 DO HTLV-1 COMO POTENCIAL ALVO PARA O DESENVOLVIMENTO DE UMA VACINA TERAPÊUTICA."

FERNANDA KHOURI BARRETO

FOLHA DE APROVAÇÃO

COMISSÃO EXAMINADORA

Dr. Lucas Pedreira de Carvalho
Pesquisador
IGM/FIOCRUZ

Dra. Isadora Cristina de Siqueira
Pesquisadora
IGM/FIOCRUZ

Dra. Dalila Luciola Zanette
Pesquisadora
IGM/FIOCRUZ

"Há verdadeiramente duas coisas diferentes: saber e crer que se sabe. A ciência consiste em saber; em crer que se sabe reside a ignorância."

Hipócrates

FONTES DE FINANCIAMENTO

Programa de Excelência em Pesquisa (PROEP-CNPq/FIOCRUZ-BA)
Programa Ciências sem Fronteiras (MEC/MCTI/CAPES/CNPq/FAPs)
Fundação Oswaldo Cruz – Instituto Gonçalo Moniz (FIOCRUZ-IGM)

BARRETO, Fernanda Khouri. Caracterização molecular da ORF-I do HTLV-1 proveniente de pacientes com diferentes perfis clínicos. 76 f. il. Tese (Doutorado em Biotecnologia em Saúde e Medicina Investigativa) – Fundação Oswaldo Cruz, Instituto Gonçalo Moniz, Salvador, 2017.

RESUMO

INTRODUÇÃO: O HTLV-1 é o agente etiológico de doenças humanas, tais como leucemia/linfoma de célula T do adulto (ATLL), paraparesia espástica tropical/mielopatia associada ao HTLV (HAM/TSP), dermatite infectiva associada ao HTLV-1 (DIH), entre outras. Estima-se que cerca de 5-10 milhões de pessoas estejam infectadas pelo HTLV-1 no mundo e apesar dessa infecção ser endêmica em diferentes regiões geográficas, ainda permanece sem métodos terapêuticos eficazes. O genoma desse retrovírus é composto por duas fitas simples de RNA, com os genes gag, pol, env e uma região próxima à extremidade 3' conhecida como pX. A região pX contém quatro quadros abertos de leitura (ORFs) que codificam proteínas regulatórias específicas. A ORF-I codifica as proteínas p12 e p8, que quando expressas em quantidades equivalentes favorecem a infectividade e persistência viral. Mutações gênicas específicas na ORF-I do HTLV-1 podem alterar o padrão de expressão e, consequentemente, a concentração relativa destas proteínas, implicando na alteração da persistência viral e no desfecho da infecção. **OBJETIVO**: Caracterizar a ORF-I do HTLV-1 de pacientes com diferentes perfis clínicos. **MATERIAL E MÉTODOS**: Inicialmente foi realizada a anotação completa do principal genoma do HTLV-1 (ATK1), utilizado como sequência referência para a caracterização molecular da ORF-I. Em seguida, 1530 sequências da ORF-I provenientes de indivíduos assintomáticos e com HAM/TSP foram submetidas a análise de *dataming* para busca de associações entre mutações, carga proviral e sintomatologia. Para avaliar o grau de conservação genética da ORF-I em outros perfis clínicos, amostras de 23 pacientes assintomáticos, 11 pacientes com DIH, 13 com ATLL e 16 com HAM/TSP foram coletadas e submetidas à amplificação e sequenciamento. As sequências foram analisadas *in silico* utilizando programas de bioinformática para caracterização molecular. **RESULTADOS**: No primeiro trabalho foram compiladas as informações sobre a posição nucleotídica inicial e final dos genes do HTLV-1 e seus respectivos produtos. Em seguida, as análises das ORF-I demonstraram que apesar da baixa diversidade genética dessa região, alterações nucleotídicas específicas quando combinadas com alta carga proviral podem estar associadas ao desenvolvimento de HAM/TSP. Os dados da análise da ORF-I proveniente de pacientes com diferentes perfis clínicos demonstraram a baixa diversidade genética desta região, corroborando com o fato de que o genoma do HTLV-1 é geneticamente estável e, portanto, o desenvolvimento de uma vacina terapêutica é viável. **CONCLUSÃO**: Esse trabalho possibilitou a disponibilização no GenBank de sequências da ORF-I do HTLV-1 provenientes de pacientes com DIH, ATLL, HAM/TSP e pacientes assintomáticos. A partir desses dados foi possível identificar a ORF-I do HTLV-1 como um possível alvo para uma vacina terapêutica baseada na capacidade de influenciar a expressão de p12 e p8.

Palavras-chave: HTLV-1; ORF-I; Mutação.

BARRETO, Fernanda Khouri. Molecular characterization of HTLV-1 ORF-I from patients with different clinical profiles. 76 f. il. (Doutorado em Biotecnologia em Saúde e Medicina Investigativa)) – Fundação Oswaldo Cruz, Instituto Gonçalo Moniz, Salvador, 2017.

ABSTRACT

INTRODUTION: The HTLV-1 is the etiological agent of some human diseases, such as adult T-cell leukaemia/lymphoma (ATLL), HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP), infective dermatitis associated to HTLV-1 (IDH), among others. It is estimated that approximately 5-10 million people are infected with HTLV-1 in the world and although this infection is endemic in different geographic regions, it still remains without effective therapeutic methods. The genome of this retrovirus is composed of two single strands of RNA, with the genes gag, pol, env and a region near the 3' end, known as pX. The pX region contains four open reading frames (ORFs) that encode specific regulatory proteins. The ORF-I encodes the p12 and p8 proteins, which when expressed in equivalent concentrations favor infectivity and viral persistence. Specific mutations in HTLV-1 ORF-I may alter the expression and, consequently, the relative concentration of these proteins, implying in viral persistence alteration and outcome of infection. **OBJECTIVE**: Characterize the HTLV-1 ORF-I from patients with different clinical profiles. **MATERIAL AND METHODS**: First, the complete annotation of the main genome of HTLV-1 (ATK1), used as a reference sequence for the molecular characterization of ORF-I, was initially performed. Then, 1530 ORF-I sequences from asymptomatic and HAM/TSP individuals were submitted to dataming analysis to search associations. To assess the ORF-I genetic conservation status in other clinical profiles, samples from 23 asymptomatic patients, 11 patients with IDH, 13 with ATLL and 16 with HAM/TSP were collected and submitted to amplification and sequencing. These sequences were analyzed in silico using bioinformatics programs for molecular characterization. between mutations, proviral load and symptomatology. **RESULTS**: In the first work, the information about the initial and final nucleotide position of the HTLV-1 genes and their respective products was compiled. Then, the ORF-I analyses demonstrated that despite the low genetic diversity of this region, specific nucleotide changes when combined with high proviral load may be associated with the development of HAM/TSP. The data of ORF-I analyses from patients with different clinical profiles demonstrated the low genetic diversity of this region, corroborating the fact that the HTLV-1 genome is genetically stable and the development of a therapeutic vaccine is viable. **CONCLUSION**: Through this work was possible provide to GenBank HTLV-1 ORF-I sequences from patients with IDH, ATLL, HAM/TSP and asymptomatic patients. From these data, it was possible to identify HTLV-1 ORF-I as a possible target for a therapeutic vaccine based on the ability to influence p12 and p8 expression.

Key-words: HTLV-1, ORF-I; Mutation.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

APC	Células apresentadoras de antígenos
ATLL	Leucemia/linfoma de células T do adulto
BRC	Receptores de células B
CA	Capsídeo
CD4+	Linfócitos T CD4+
CD8+	Linfócitos T CD8+
CTL	Linfócito T citotóxico
DC	Célula dendrítica
DIH	Dermatite infectiva associada ao HTLV-1
DNA	Ácido desoxirribonucleico
Env	Envelope
gag	Grupo antigênico
GLUT-1	Molécula transportadora de glicose tipo 1
HAM/TSP	Paraparesia Espástica Tropical/ Mielopatia Associada ao HTLV
HBZ	HTLV-1 bZIP factor gene
HIV	Vírus da Imunodeficiência Humana
HLA	Antígeno leucocitário humano
HSPG	Heparan Sulfato-Proteoglicano
HTLV	Vírus linfotrópico de células T humanas
HTLV-1	Vírus linfotrópico de células T humanas tipo 1
HTLV-2	Vírus linfotrópico de células T humanas tipo 2
HTLV-3	Vírus linfotrópico de células T humanas tipo 3
HTLV-4	Vírus linfotrópico de células T humanas tipo 4
IFN-γ	Interferon-γ
IL	Interleucinas
IN	Integrase
kb	Kilobases
LFA-1	Antígeno-1 associado à função leucocitária
LTR	Repetições terminais longas (*long terminal repeat*)
LZ	*Leucine zíper*
MA	Matriz viral
MHC-I	Complexo de Histocompatibilidade principal classe 1

MHC-II	Complexo de Histocompatibilidade principal classe 2
NC	Nucleocapsídeo
NCI-NIH	*National Institutes of Health*
NFAT	Fator nuclear de ativação de células T
NK	Células *Natural Killer*
NRP-1	Neurofilina 1
OMS	Organização Mundial de Saúde
ORF	Fase de leitura aberta (*open reading frame*)
pb	Pares de bases
PCR	Reação em cadeia da polimerase
pol	Polimerase
PTLV	Vírus linfotrópico de células T de primatas
RNA	Ácido ribonucleico
RNAm	Ácido ribonucleico mensageiro
RT	Transcriptase reversa
SH3	*Proline-rich Src homology*
STLV	Vírus linfotrópico de células T em Símios
SU	Proteína de superfície
TCR	Receptores de células T
Th1	Resposta T *helper* do tipo 1
Th2	Resposta T *helper* do tipo 2
TM	Proteína transmembrana
TNF	Fator de necrose tumoral
TRE	Elementos Responsivos de Tax
WB	Western Blot

SUMÁRIO

1 INTRODUÇÃO

O HTLV-1 é um retrovírus oncogênico, responsável pelo desenvolvimento da Leucemia/Linfoma de Células T do Adulto (ATLL) e também pelo desenvolvimento de condições inflamatórias, como a Paraparesia Espástica Tropical/Mielopatia associada ao HTLV-1 (HAM/TSP), dermatite infectiva associada ao HTLV-1 (DIH), uveíte, síndrome de Sjögren, broncoalveolíte, artrite, entre outros. Sabe-se que a maioria dos pacientes infectados pelo HTLV-1 não desenvolvem manifestações clínicas, mas, dentre os pacientes sintomáticos, a ATLL, HAM/TSP e DIH são detectadas com maior prevalência (GESSAIN, A et al., 1985; OSAME et al., 1986; MORGAN et al., 1989; NISHIOKA et al., 1989; EGUCHI et al., 1992; MOCHIZUIKI et al., 1992).

Diversos trabalhos vêm sendo realizados a fim de se desenvolver terapias específicas contra a infecção pelo HTLV-1, uma vez que os pacientes assintomáticos permanecem sem tratamento e os pacientes que apresentam manifestações clínicas são tratados de maneira paliativa. O fato de não haver uma terapia específica e eficaz contra a infecção pelo HTLV-1 compromete a erradicação deste vírus. Além disso, a ausência de um método de controle para a carga proviral impacta diretamente no aumento do risco de transmissão viral e no desenvolvimento de doenças associadas, como a HAM/TSP.

Estudos demonstram que a persistência da infecção pelo HTLV-1 está relacionada à expressão balanceada das proteínas p12 e p8, codificadas a partir da ORF-I da região gênica *pX*. Esses dados sugerem que mutações na ORF-I alteram a expressão dessas proteínas e que concentrações equivalentes de ambas são necessárias para impedir o reconhecimento e lise das células infectadas pelas células T citotóxicas (CTL). Estudos demonstram ainda que a p12 diminui a expressão do complexo de histocompatibilidade tipo I (MHC-I) provocando a inibição da apresentação dos peptídeos virais para as CTLs, além de diminuir também a expressão de ICAM-1 e ICAM-2 nas células T, inibindo assim a adesão e o reconhecimento das células T infetadas pelas células *Natural Killer* (NK). Da mesma forma, a p8 também influencia o curso da infecção, já que está associada com a disseminação viral através da formação de conduítes (JOHNSON et al., 2001; BANERJEE et al., 2007; PISE-MASISON et al., 2014).

Considerando a influência da expressão balanceada da ORF-I do HTLV-1 para a infectividade e persistência viral, este estudo se propõe a avaliar se essa região poderia ser utilizada como alvo para o desenvolvimento de uma vacina terapêutica capaz de

alterar o equilíbrio da expressão de p12 e p8 (PISE-MASISON et al., 2014). A hipótese central da presente tese é que a ORF-I do HTLV-1 é conservada entre os diferentes perfis clínicos e, portanto, uma vacina de DNA capazes de competir pelos sítios de clivagem da p12 poderia impedir a formação de p8 e favorecer o reconhecimento das células infectadas pelas CTLs. A partir dos resultados obtidos estudos futuros poderão ser conduzidos com o objetivo de desenvolver essa vacina.

Adicionalmente, os dados obtidos poderão ser utilizados pela comunidade científica para análises *in silico* do genoma do HTLV-1, visto que esta tese disponibiliza a anotação completa do clone molecular do HTLV-1 (ATK1), além de sequências da ORF-I provenientes de pacientes com DIH, ATLL, HAM/TSP e assintomáticos.

2 OBJETIVOS

2.1 Geral

Caracterizar a ORF-I do HTLV-1 proveniente de paciente com diferentes perfis clínicos.

2.2 Específicos

- Realizar a anotação completa do principal clone molecular do HTLV-1;
- Investigar possíveis correlações entre mutações na ORF-I do HTLV-1, carga proviral e desenvolvimento de HAM/TSP;
- Avaliar a diversidade genética da ORF-I do HTLV-1 proveniente de pacientes com ATLL, DIH, HAM/TSP e assintomáticos.

3 REFERENCIAL TEÓRICO

3.1 VÍRUS LINFOTRÓPICO DE CÉLULAS T HUMANAS (HTLV)

O vírus linfotrópico de células T humanas (HTLV) foi o primeiro retrovírus humano identificado, a partir de um paciente com linfoma cutâneo de células T, nos Estados Unidos da América em 1979 (POIESZ et al., 1980). Na mesma época, pesquisadores japoneses isolaram e caracterizaram o mesmo vírus, que passou a ser classificado como vírus linfotrópico de células T humanas do tipo 1 (HTLV-1) (HINUMA et al., 1981). Posteriormente, em 1982, foi identificado o vírus linfotrópico de células T humanas do tipo 2 (HTLV-2), numa linhagem imortalizada de células T obtidas de um paciente com tricoleucemia (KALYANARAMAN et al., 1982). Em 2005, foram identificados o vírus linfotrópico de células T humanas do tipo 3 (HTLV-3) e vírus linfotrópico de células T humanas do tipo 4 (HTLV-4), em populações do sul de Camarões (WOLFE et al., 2005).

O HTLV pertence à família *Retroviridae*, à subfamília *Orthoretrovirinae* e ao gênero *Deltaretrovirus*. Estudos sugerem que esse retrovírus pode ter emergido na África, a partir do contato entre humanos e primatas não-humanos infectados com o vírus linfotrópico de células T de símios (STLV). A África é o único continente onde todos os vírus linfotrópicos de células T de primatas (PTLV) foram isolados, e acredita-se que o HTLV foi levado ao novo mundo (Caribe, Estados Unidos e América do Sul) pelos negros africanos durante o período de tráfico de escravos no século XVI (COURGNAUD et al., 2004; SANTOS e LIMA, 2005; VERDONCK et al., 2007).

3.2 DISTRIBUIÇÃO DO HTLV-1

Dados epidemiológicos sugerem que a infecção pelo HTLV-1 tem distribuição mundial, no entanto, algumas regiões são consideradas áreas endêmicas, como o sudoeste do Japão (YAMAGUCHI, 1994; MUELLER et al., 1996), alguns países no Caribe

(HANCHARD et al., 1990), África sub-Saariana (GESSAIN, A. e DE THE, 1996) e áreas localizadas no Irã e Melanésia (MUELLER, 1991) (Figura 1). O Japão e a África exemplificam a não homogeneidade da distribuição, e são considerados os continentes com maior número de indivíduos infectados pelo HTLV-1, seguido da América do Sul (GESSAIN, ANTOINE e CASSAR, 2012).

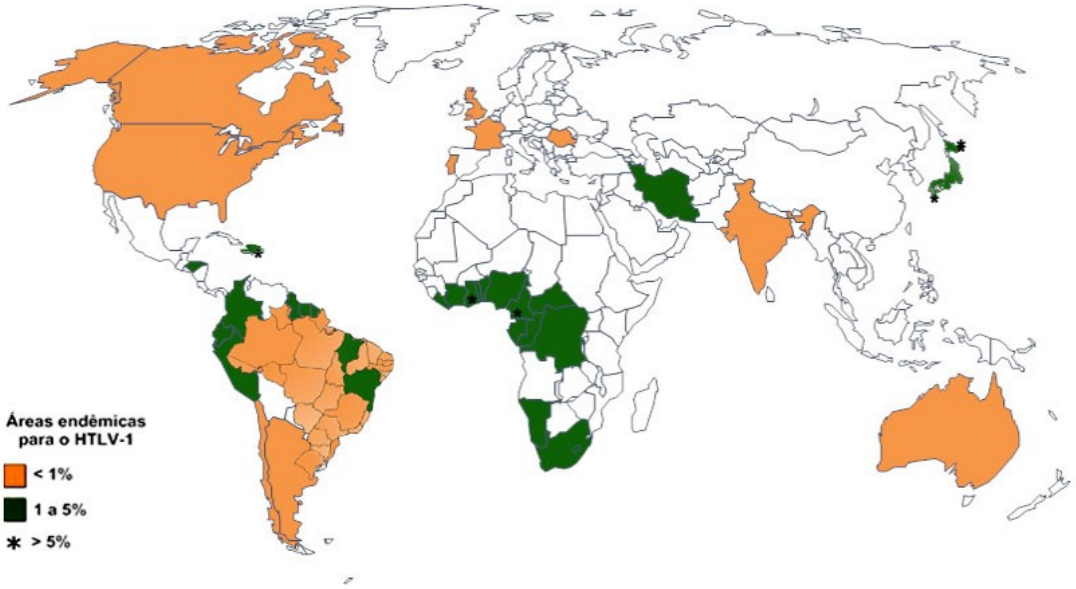


Figura 1. Distribuição do HTLV-1 no mundo. Adaptado de PROIETTI, 2006.

Estima-se que aproximadamente 5-10 milhões de pessoas estejam infectadas pelo HTLV-1 no mundo (GESSAIN, ANTOINE e CASSAR, 2012). No Brasil, a infecção pelo HTLV atinge, aproximadamente, 2,5 milhões de pessoas (CARNEIRO-PROIETTI et al., 2002; CATALAN-SOARES et al., 2004) e apesar da escassez dos dados epidemiológicos, também pode-se observar o fenômeno de distribuição heterogênea da infecção, sendo as maiores prevalências observadas em doadores de sangue nos estados do Maranhão, Bahia, Pará e Pernambuco (CATALAN-SOARES et al., 2005) (Figura 2). Considerando a população total, Salvador apresenta uma prevalência de 1,78% de indivíduos infectados pelo HTLV-1 (DOURADO et al., 2003).

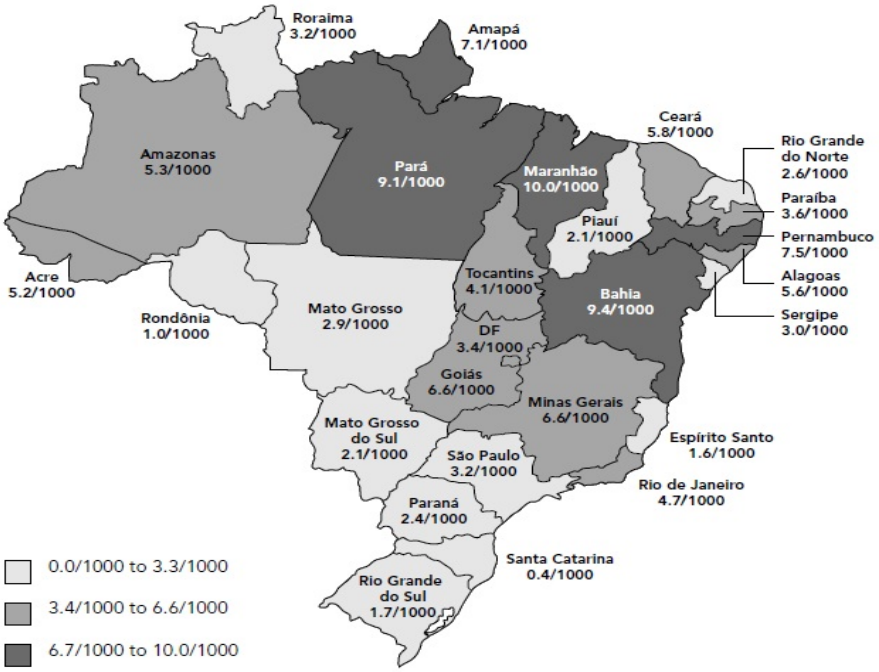


Figura 2. Mapa da soroprevalência para o HTLV em doadores de sangue no Brasil. Adaptado de CATALAN-SOARES et al., 2005.

3.3 ESTRUTURA DO HTLV-1

O HTLV-1 possui uma estrutura morfológica similar a de outros retrovírus. É constituída, basicamente, por um envelope, uma matriz protéica e um nucleocapsídeo. O envelope representa a superfície do vírion e é composto por duas subunidades protéicas glicosiladas: a proteína de superfície (SU), denominada gp46, e a proteína transmembrana (TM), denominada gp21, que ancora a gp46. Logo após o envelope, encontra-se a matriz viral (MA), seguida do capsídeo (CA), que apresenta simetria icosaédrica e é composto principalmente por proteínas codificadas a partir do gene *gag*. No interior do capsídeo encontra-se o genoma viral, composto por duas fitas simples de RNA senso positivo, associadas às proteínas do nucleocapsídeo (NC), além de outras proteínas, como a transcriptase reversa (RT) e a integrase (IN) (Figura 3).

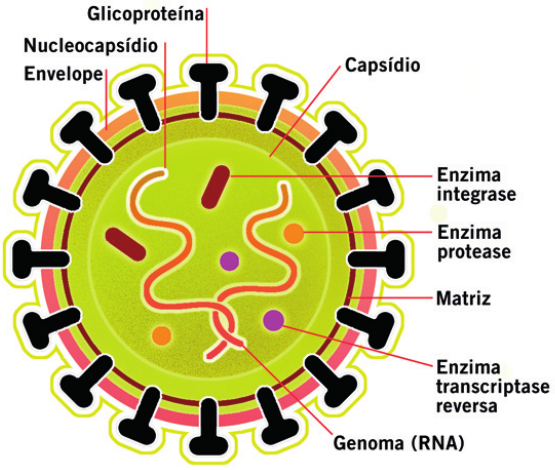


Figura 3. Estrutura morfológica do HTLV-1. Desenho esquemático adaptado de www.researchgate.net/publication/264543722Opapeldosvirusnaarvoredavida.

3.4 ORGANIZAÇÃO GENÔMICA DO HTLV-1

O genoma viral localiza-se no interior do capsídeo e possui 9032 pb (SEIKI et al., 1983). É composto por duas fitas simples de RNA, que codificam os genes *gag* (grupo antigênico), *pro* (protease), *pol* (polimerase) e *env* (envelope), além de uma sequência próxima à extremidade 3', conhecida como região *pX*. Similar a outros retrovírus, o RNA mensageiro (RNAm) transcrito a partir do gene *gag* codifica a proteína p55 que, posteriormente, é clivada em p19, p24 e p15. Já o gene *pol* é responsável por codificar a transcriptase reversa e integrase, enquanto a protease é expressa a partir do gene *pro*. Da mesma forma, o RNAm transcrito a partir do gene *env* codifica a proteína p63, que após a clivagem origina as proteínas do envelope gp46 e gp21. Em cada extremidade do genoma existem sequências de repetições terminais longas (LTR – *long terminal repeat*), que são essenciais na integração do DNA proviral ao DNA do hospedeiro e também para a regulação transcricional do genoma do vírus (Figura 4) (GREEN e CHEN, 2001).

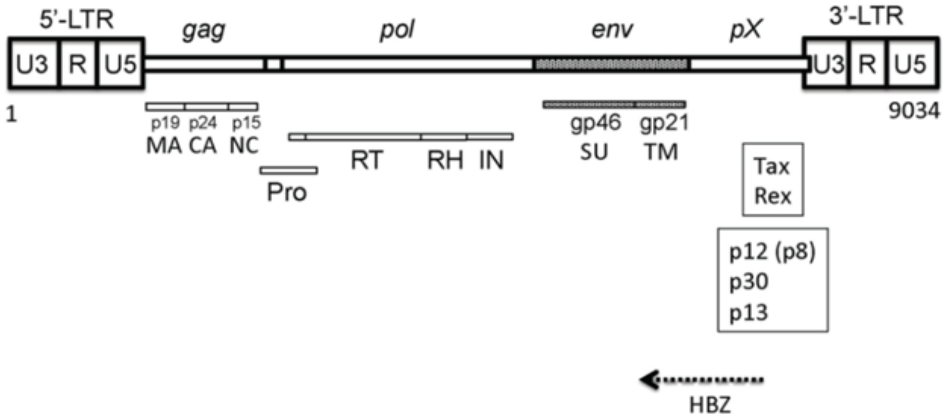


Figura 4. Estrutura genômica do HTLV-1 e processamentos de RNA. Desenho esquemático adaptado de PROIETTI, 2006; HOSHINO, 2012.

3.4.1 Região promotora

A região LTR do HTLV-1 contém elementos virais promotores e regulatórios e exerce um papel fundamental na integração do genoma viral no genoma da célula hospedeira (GREEN e CHEN, 2001). Essa região divide-se em três subunidades: U3, R e U5. A região U3 possui elementos necessários para o controle da transcrição viral, como sítios de poliadenilação, motivo TATA box, e três repetições de 21 pares de base (pb) conhecidas como Elementos Responsivos de Tax (TRE) (BRADY et al., 1987). Vale ressaltar que a proteína Tax não se liga especificamente a região TRE, mas interage com fatores de transcrição, como CREB/ATF, que reconhecem o motivo CRE, localizado nessa região. A formação desse complexo possibilita o recrutamento de fatores co-ativadores p300/CBP, facilitando a estabilidade da ligação e consequentemente, estimulado à transcrição viral (BARBEAU e MESNARD, 2011).

Tem sido demonstrado que a transcrição viral é controlada pela ação de proteínas específicas, denominadas fatores de transcrição, como CREB/ATF, NF-Kb, e SRF, que se ligam especificamente a motivos presentes na região promotora. Da mesma forma, processos epigenéticos, como a metilação da região LTR também influenciam a expressão gênica. Trabalhos sugerem que a região LTR 5’ é hipermetilada em relação à LTR 3’, tanto para o HTLV-1 quanto para o HTLV-2. Devido a isso, a expressão gênica regulada pela LTR 3’ mantém-se constante, ao contrário da expressão associada a LTR 5’. Logo, as proteínas HBZ (HTLV-1) e APH-2 (HTLV-2), que são controladas pela

região LTR 3', possuem expressão contínua (GAUDRAY et al., 2002; HALIN et al., 2009).

A transcrição a partir da região LTR 5' gera uma molécula de RNA completa (9 kb), utilizada para a síntese dos produtos dos genes *gag*, *pro*, e *pol*. Um RNA subgenômico (4 kb) é sintetizado a partir de uma etapa única de processamento da molécula de RNA completa, e codifica as glicoproteínas do envelope viral (*env*). Da mesma forma, outra molécula de RNA (2 kb) é gerada após a remoção de dois *íntrons,* e codifica as proteínas regulatórias da região *pX* (RENDE et al., 2012).

3.4.2 Genes estruturais

O gene *gag* corresponde aos nucleotídeos 824 a 2113 no genoma do HTLV-1, e essa região é inicialmente traduzida como um precursor (p55), que após a clivagem origina as proteínas da matriz (p19), proteínas do capsídeo (p24), e a proteína do nucleocapsídeo (p15). A extremidade 3' do gene *gag* sobrepõe-se ao início do gene *pro,* que compreende a posição 1960 a 2778 no genoma do HTLV-1. O gene *pro* codifica a enzima protease (p14), que atua no processo de clivagem das cadeias polipeptídicas originando proteínas maduras (JOHNSON et al., 2001).

O gene *pol,* por sua vez, ainda não possui localização inicial descrita na literatura, mas sabe-se que é o gene responsável por codificar a transcriptase reversa (TR), que atua na síntese do DNA viral e a integrase (IN), fundamental no processo de integração do DNA viral ao genoma da célula hospedeira (PROIETTI, 2006).

O gene *env* corresponde a sequência de nucleotídeos localizada na posição 5203 a 6669 no genoma do HTLV-1, e assim como o gene *gag,* codifica inicialmente um precursor. Esse precursor (gp63) é clivado entre a posição 6138-6139 para gerar uma glicoproteína transmembrana (gp21) e uma glicoproteína de superfície (gp46). Assim como em outros retrovírus, a glicoproteína transmembrana ancora, através de associações não covalentes, a glicoproteína de superfície na bicamada lipídica do envelope. Dessa forma, a gp21 e gp46 ficam expostas ao sistema imune e, consequentemente, estão associadas à resposta humoral e celular (DELAMARRE et al., 1996; MANEL, 2005).

3.4.3 Genes regulatórios

A região *pX* do genoma do HTLV-1 é responsável por codificar proteínas regulatórias e acessórias importantes durante os mecanismos de replicação viral. A transcrição dessa região resulta em formas alternativas de RNA mensageiro, que possuem quatro fases de leitura aberta (ORFs), conforme figura 5.

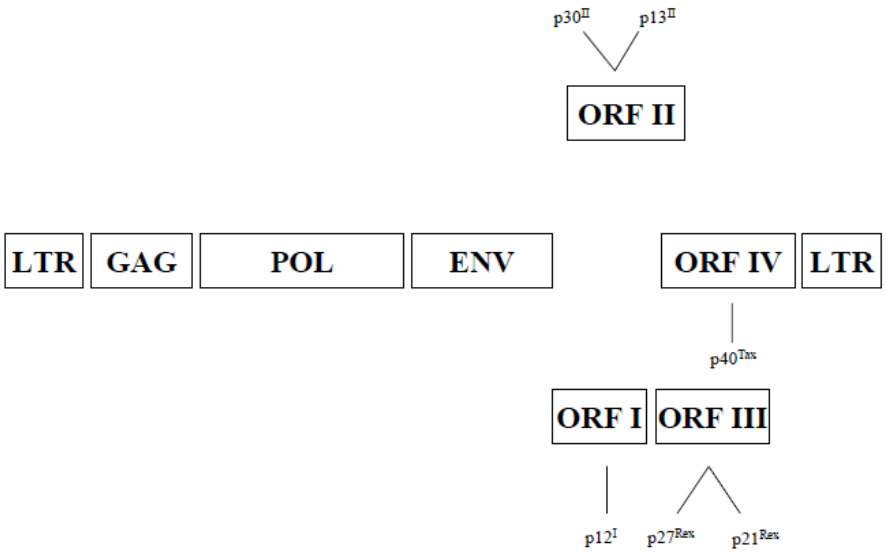


Figura 5. Representação esquemática da região pX do genoma do HTLV-1. Adaptado de JOHNSON et al., 2001.

A ORF-I codifica a proteína p12, que reside no retículo endoplasmático e é capaz de aumentar a concentração de cálcio intracelular via receptores de trifosfato de inositol (ALBRECHT et al., 2002; DING et al., 2002).

O aumento da concentração de cálcio pode influenciar diversos mecanismos celulares, como a regulação da transcrição viral através da modulação do cofator p300 e a estimulação da propagação viral através do contato célula-célula, por meio da indução do agrupamento do antígeno-1 associado à função leucocitária (LFA-1) (KIM et al., 2006; NAIR et al., 2006). Além disso, o acúmulo do cálcio no citoplasma ativa a calmodulina e calcineurina, que através de desfosforilação ativa o fator nuclear de ativação de células T (NFAT). O NFAT é então transportado para o núcleo e promove a ativação, diferenciação e proliferação dos linfócitos T (FUKUMOTO et al., 2007). De maneira

oposta, a p12 também é capaz de inibir a proliferação de linfócitos T, já que possui um domínio para a calcineurina e pode, portanto, competir com o NFAT pelo sítio de ligação (KIM et al., 2003; FUKUMOTO et al., 2007).

Sabe-se também que a p12 diminui a expressão de moléculas do complexo de histocompatibilidade tipo I (MHC I), já que é capaz de se ligar ao MHC-I recém-sintetizado, impedindo a sua maturação (JOHNSON et al., 2001). As moléculas de MHC-I não maturadas são degradadas pelo proteassoma e as células que apresentam uma redução da expressão de MHC-I são naturalmente identificadas e lisadas pelas células *Natural Killer* (NK). Entretanto, a p12 também altera a expressão de moléculas de adesão ICAM-1 e ICAM-2, dificultando a interação entre as células e o reconhecimento da célula infectada (BANERJEE et al., 2007).

Estudos demonstram que a p12 interage com a ATPase presente na membrana dos lisossomos e responsável pelo transporte de íons H^+. A internalização adequada desses prótons é fundamental para a correta acidificação do meio e, consequentemente, funcionamento das hidrolases ácidas (KORALNIK et al., 1995; JONES, K. S. et al., 2008).

A p12 está susceptível a duas clivagens protéicas. A primeira ocorre entre os aminoácidos 9 e 10 e resulta na perda do sinal de retenção para o retículo endoplasmático. A segunda clivagem ocorre entre os aminoácidos 29 e 30 e dá origem a proteína p8. Após a clivagem, a p8 se desloca para a superfície da célula infectada e atua na disseminação viral através da formação de conduítes (VAN PROOYEN et al., 2010) (Figura 6).

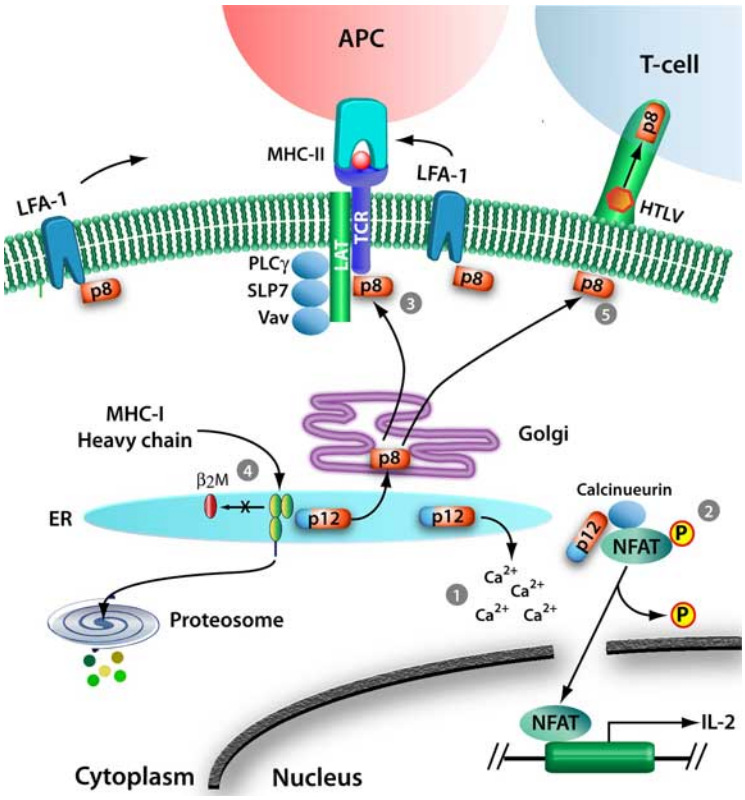


Figura 6. Funções das proteínas p12 e p8, codificadas a partir da ORF-I do HTLV-1. Desenho esquemático adaptado de EDWARDS et al., 2011.

As proteínas acessórias p13 e p30, codificadas a partir da ORF II contribuem para a manutenção da carga proviral, ativação das células do hospedeiro e regulação da transcrição gênica (BINDHU et al., 2004).

Já a ORF III codifica uma importante proteína regulatória, denominada *Rex* (p27). Essa proteína é considerada essencial para a regulação da expressão gênica viral por meio de eventos pós-traducionais. *Rex* é indispensável para a eficiente multiplicação, infecção e disseminação viral, além de regular a indução das diferentes fases do ciclo celular do HTLV-1 (YOUNIS e GREEN, 2005).

A ORF IV, por sua vez, codifica a proteína regulatória viral *Tax* (p40), essencial para a replicação viral e para a transformação celular (PROIETTI, 2006). *Tax* pode ser encontrada tanto no núcleo como no citoplasma da célula e, além de regular a transcrição viral, pode aumentar a expressão de citocinas e receptores envolvidos no crescimento e proliferação celular (SMITH e GREENE, 1992; TSCHACHLER et al., 1993). Além disso, *Tax* é capaz de reprimir a ação de genes celulares envolvidos na inibição do crescimento celular, bem como inibir o reparo do DNA e os eventos de apoptose

(FRANCHINI, 1995). Essa proteína atua como agente principal no desenvolvimento de diversas patologias associadas ao HTLV-1.

Recentemente, foi identificado na região *pX* o gene *HTLV-1 bZIP factor gene* (HBZ), codificado sob o controle da região LTR 3' (MATSOUKA e GREEN, 2009). Estudos sugerem que o HBZ pode atuar tanto como molécula de RNA, quanto em sua forma protéica, e tem como principal função a regulação da transcrição viral, e o aumento da proliferação das células T (MESNARD et al., 2006). Além disso, a proteína HBZ inibe a transativação da transcrição viral mediada por *Tax*, por se heterodimerizar com fatores de transcrição da família CREB (MATSOUKA, 2005).

3.5 CICLO DE DISSEMINAÇÃO DO HTLV-1

A replicação clássica dos retrovírus ocorre, basicamente, em nove etapas: adsorção do vírus; fusão da membrana da célula hospedeira com o envelope viral; internalização da partícula viral; transcrição reversa do genoma viral; inserção do DNA linear no genoma da célula hospedeira; síntese de RNA viral; tradução das proteínas virais; montagem do capsídeo e brotamento (Figura 7).

A primeira etapa do ciclo clássico de replicação consiste na ligação da glicoproteína de superfície gp46 a moléculas na superfície celular, incluindo proteoglicanos heparan-sulfato, neuropilina-1 e um transportador de glicose de classe 1 (GLUT-1). Essa interação expõe a glicoproteína transmembrana gp21 e isso possibilita a fusão das membranas viral e celular. Após a internalização da partícula viral, o RNA é transcrito em DNA pela ação da transcriptase reversa e se insere no genoma da célula hospedeira, passando a ser denominado DNA proviral (PINON et al., 2003; MANEL, 2005; GHEZ et al., 2006; JONES, KATHRYN S. et al., 2011).

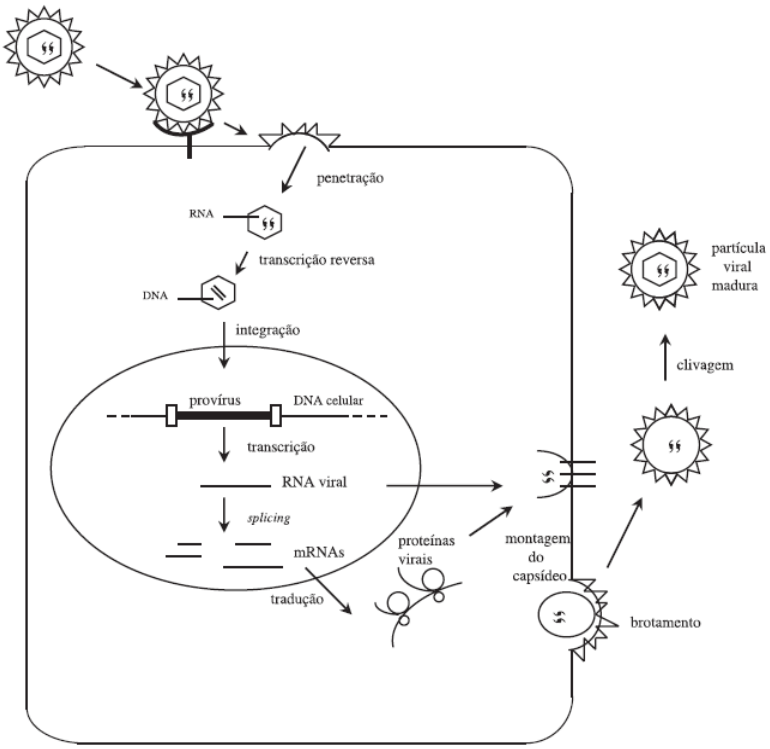


Figura 7. Ciclo de replicação clássica dos retrovírus PROIETTI, 2006.

Ao contrário de outros retrovírus, a replicação do HTLV-1 não ocorre somente via mecanismo clássico. Diversas formas de disseminação já foram associadas a esse retrovírus, onde se observa a presença de partículas virais induzindo o contato entre uma célula infectada e uma célula não infectada.

Uma vez que a infecção ocorre principalmente pelo contato célula-célula, as novas partículas virais são capazes de induzir eventos de polarização das células, para facilitar a passagem de proteínas e do genoma viral, num fenômeno conhecido como sinapse virológica (BANGHAM, 2003). Esse processo ocorre quando uma célula infectada entra em contato com outra célula não infectada, induzindo a formação de um canal protéico na junção célula-célula (Figura 8). A formação desta estrutura permite o acúmulo de proteínas codificadas a partir do gene *gag* e de material genômico na interface da sinapse, culminando com a passagem desse material para a célula não infectada (MATSUOKA e JEANG, 2007).

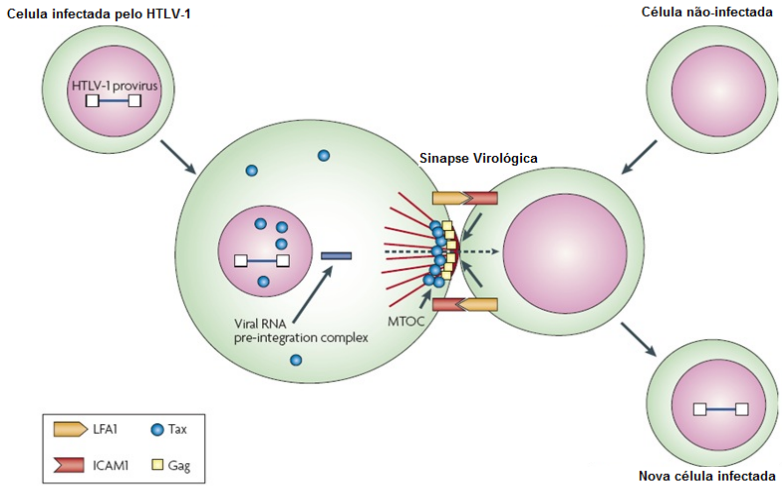
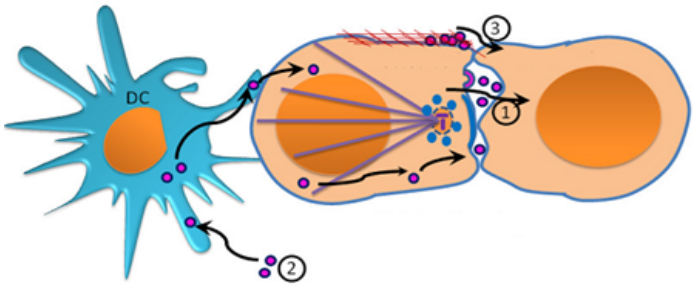


Figura 8. Sinapse Viral. Adaptado de MATSUOKA e JEANG, 2007.

Além da disseminação viral através do modelo clássico e da sinapse virológica, recentemente o HTLV-1 foi detectado em estruturas ricas em carboidratos, semelhantes a biofilmes, presentes na superfície de células infectadas. Essas estruturas favorecem o contato entre linfócitos infectados e linfócitos não-infectados, facilitando a transmissão viral (PAIS-CORREIA et al., 2009) (Figura 9).



1) Sinapse viral
2) Modelo clássico de replicação dos retrovírus
3) Formação de biofilmes

Figura 9. Disseminação viral através da formação de biofilmes. DC: Células Dendríticas. Adaptado de NEJMEDDINE e BANGHAM, 2010.

Em 2010, Nancy Van Prooyen e colaboradores detectaram proteínas virais, como p8 e proteínas traduzidas a partir dos genes *env* e *gag*, em conduítes formados entre linfócitos infectados e linfócitos não-infectados, sugerindo, portanto, um novo modelo de disseminação viral utilizado pelo HTLV-1 (Figura 10).

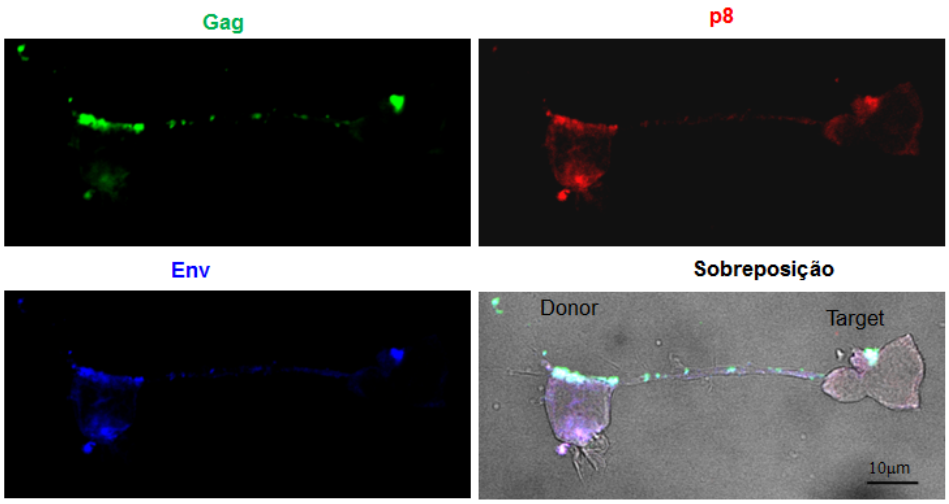


Figura 10. Formação de Conduítes. Adaptado de VAN PROOYEN et al., 2010.

Apesar de já terem sido descritas diversas formas de disseminação viral, o HTLV-1 utiliza principalmente a expansão clonal das células infectadas, via mitose (WATTEL et al., 1995; CIMARELLI et al., 1996). Logo, esse retrovírus não utiliza com frequência a transcriptase reversa, descrita na literatura como enzima com altas taxas de erros (PRESTON e DOUGHERTY, 1996). Esses erros resultam em mutações que diminuem a estabilidade do genoma viral. Portanto, o HTLV-1 é um retrovírus estável, quando comparado com outros retrovírus, como o Vírus da Imunodeficiência Humana (HIV), por exemplo.

3.6 MANIFESTAÇÕES CLÍNICAS E FISIOPATOLOGIA

O HTLV-1 foi inicialmente associado a leucemia/linfoma de células T do adulto (ATLL) (YOSHIDA et al., 1982). Posteriormente, a infecção pelo HTLV-1 foi associada as patologias neurológicas paraparesia espástica tropical (TSP) e mielopatia associada ao HTLV (HAM), as quais, em 1988, foram admitidas por Román e Osame como manifestações complementares, compondo uma síndrome neurológica hoje conhecida como HAM/TSP (GESSAIN, A. et al., 1985).

A ATLL é um linfoma/leucemia agressivo, que tem seu curso clínico classificado em quatro estágios: aguda, crônica, linfoma e *smouldering* (SHIMOYAMA, 1991). Além desses estágios clínicos, existe ainda outra categoria conhecida como ATLL cutânea, cujas manifestações são restritas à pele, e que posteriormente foi subdividida em duas outras classificações: tumoral e eritematosa (BITTENCOURT et al., 2007). A ATLL caracteriza-se pela infiltração de células T $CD4^+$ nos linfonodos, baço, trato gastro-intestinal e pele, além da presença de células T anormais com núcleo lobulado ou em forma de flor (*flower cells*) (MATSOUKA, 2005). A proporção de ocorrência de ATLL em homens e mulheres infectados no Japão é de seis e dois por cento, respectivamente (YASUNAGA e MATSOUKA, 2007).

A HAM/TSP, por sua vez, caracteriza-se por ser uma mielopatia lentamente progressiva, que tem como consequência a paraparesia dos membros inferiores, espasticidade e hiperreflexia (ROMAN e OSAME, 1988). Esta patologia, considerada uma doença de alta morbidade, está presente em um a cinco por cento dos pacientes infectados pelo vírus e atinge com maior frequência as mulheres (KAPLAN et al., 1990). Apesar de ainda não ser completamente esclarecida, existem três hipóteses que buscam explicar a patogênese da HAM/TSP. A primeira hipótese sugere que linfócitos T $CD8^+$ específicos anti-HTLV-1 atravessariam a barreira hematoencefálica e destruiríam as células da glia infectadas pelo HTLV-1, através de mecanismos de citotoxicidade direta (NAKAMURA et al., 1993). A segunda hipótese envolve um processo de mimetismo molecular entre o antígeno Tax e uma proteína neuronal, que acarretaria em um processo inflamatório autoimune com consequente lesão neuronal (LEVIN et al., 2002).

A terceira, e mais aceita hipótese, propõe que a patogênese da HAM/TSP esteja relacionada à desmielinização local atribuída à invasão de linfócitos infectados no sistema nervoso central, e o desencadeamento de uma resposta inflamatória crônica com destruição tecidual. Esse processo ocorreria a partir de células T $CD4^+$ infectadas com o HTLV-1 e linfócitos T $CD8^+$ específicos anti-Tax que migrariam através da barreira hematoencefálica e provocariam a destruição das células da glia pelas citocinas liberadas pelos CTLs contra as células T $CD4^+$ infectadas (OSAME, 2002).

Já foi relatado que as células infectadas pelo HTLV-1 são encontradas em uma frequência duas vezes maior no líquor quando comparado com o sangue periférico (HAYASHI et al., 2008). Vale ressaltar que ainda não foi demonstrada a infecção de células do sistema nervoso central pelo HTLV-1, mas os linfócitos T $CD8^+$ específicos anti-HTLV-1, que atravessam a barreira hematoencefálica, secretam citocinas

neurotóxicas, como interferon-γ (IFN-γ) e fator de necrose tumoral (TNF) (YAMANO et al., 2002). Em pacientes com HAM/TSP observa-se a predominância da resposta T *helper* do tipo 1 (Th1) e a redução de citocinas T *helper* do tipo 2 (Th2), como a interleucina-4 (IL-4), mas apesar do aumento da produção de IFN-γ, ainda não foi comprovada a associação entre os níveis dessa citocina e a gravidade da doença (MUNIZ et al., 2006; MONTANHEIRO et al., 2009).

Da mesma forma, a alta carga proviral resultante da expansão clonal de linfócitos T $CD4^+$ infectados pelo HTLV-1 leva a um aumento na expressão de antígenos virais e, consequentemente, aumenta a ativação de CTLs específicas. Nesse sentido a carga proviral é o único biomarcador de progressão para HAM/TSP (JACOBSON, 2002; GONCALVES et al., 2008).

O HTLV-1 também está associado a outras patologias, como dermatite infectiva associada ao HTLV-1 (GONÇALVES et al., 2003), uveíte associada ao HTLV (HAU) (MOCHIZUIKI et al., 1992), além de doenças reumáticas, como síndrome seca e artrite reumatóide (NISHIOKA, 1996; MCCALLUM et al., 1997).

A infecção pelo HTLV-1 não necessariamente implica no desenvolvimento de processos patológicos e ainda não se sabe o que determina a manifestação da doença e a alta morbidade relacionada. Acredita-se que fatores relacionados ao hospedeiro, bem como variações genéticas virais podem influenciar o desfecho da infecção.

Estudos apontam que a carga proviral e a resposta imune normalmente apresentam-se elevadas em indivíduos sintomáticos, quando comparados com indivíduos assintomáticos. Tem sido sugerido que essa resposta imunológica é influenciada por fatores genéticos individuais como polimorfismos em genes de HLA (*Human Lecocitary Antigen*) e genes envolvidos na resposta imune (PROIETTI, 2006). Além disso, a análise de indivíduos com diferentes estágios clínicos, mas apresentando níveis semelhantes de carga proviral, demonstrou que células de pacientes assintomáticos produzem menores níveis das citocinas pró-inflamatórias: TNF e IFN-γ (NISHIMURA et al., 2000). Isto sugere que esta baixa produção seria importante para a manutenção do estado assintomático e que outros fatores, além da carga proviral, também devem influenciar a sintomatologia da infecção (FURUKAWA et al., 2003).

De acordo com pesquisas recentes, variações no genoma do HTLV-1 também podem alterar o curso da infecção. Análises de sequências do gene *env* demonstram a presença de mutações na região gênica que codifica a gp46 e gp21. Miranda e colaboradores descreveram cinco mutações na gp46 (F14S, S35L, N42H, G72S e V247I)

e uma mutação na gp21 (Y477H) e sugerem que essas mutações levam a uma possível mudança no perfil físico-químico dessas glicoproteínas (MOTA-MIRANDA et al., 2013a; MOTA-MIRANDA et al., 2013b). Da mesma forma, também já foram descritas mutações em domínios de ativação do HBZ, além de mutações na ORF-I que podem influenciar a persistência e infectividade viral, bem como o desfecho da infecção (MOTA-MIRANDA et al., 2013b; PISE-MASISON et al., 2014). Ao realizar análises na ORF-I, Pise-Masison e colaboradores identificaram mutações responsáveis pela alteração da concentração intracelular relativa das proteínas p12 e p8, e demonstraram a importante dependência da expressão de ambas as proteínas para a infectividade e persistência viral (PISE-MASISON et al., 2014).

3.7 RESPOSTA IMUNE NA INFECÇÃO PELO HTLV-1

A resposta imune contra a infecção pelo HTLV-1 pode ser categorizada em dois grandes grupos: resposta imune inespecífica (inata) e específica (adaptativa). A resposta imune inata ocorre de maneira rápida, a partir da ação de neutrófilos e monócitos, ativação de células *Natural Killer* (NK), bem como ativação do sistema complemento (PROIETTI, 2006). Embora os mecanismos da atividade inata sejam de extrema importância no momento inicial da infecção, é a resposta imune adaptativa que exerce o papel mais relevante no controle das infecções crônicas, como a infecção pelo HTLV-1 (LAL et al., 1992).

Durante a infecção viral, os linfócitos B interagem com proteínas do HTLV-1, através dos receptores de células B (BRC), também denominados imunoglobulinas de membrana, internalizando as estruturas antigênicas. Em seguida, esses linfócitos B transformam-se em plasmócitos e secretam imunoglobulinas específicas anti-HTLV-1, com o intuito de neutralizar ou opsonizar o vírus (PROIETTI, 2006).

No âmbito da resposta imune humoral, o gene *env* tem fundamental importância, já que a maioria dos anticorpos produzidos em resposta à infecção natural pelo HTLV-1 inclui aqueles direcionados contra as glicoproteínas do envelope. Sabe-se inclusive que a região central da gp46 é responsável por induzir 90% dos anticorpos anti-*env* que são produzidos nos indivíduos infectados (SHERMAN et al., 1993). Isso ocorre porque após

a infecção as proteínas do envelope são expressas na superfície das células infectadas e são, portanto, reconhecidas pelo hospedeiro como alvos da resposta imune.

Indivíduos infectados pelo HTLV-1 desenvolvem uma resposta imune significativa contra os produtos proteicos do gene *env*, mas apesar dessa elevada resposta, os primeiros anticorpos específicos anti-HTLV-1 detectados são anti-*gag* e, portanto, outras proteínas também podem ser utilizadas como marcadores da infecção, tais como a proteína p19 e p24. Trabalhos recentes têm demonstrado o aumento na expressão do antígeno viral p19 com o aumento da proliferação das células infectadas (XIE e GREEN, 2005).

Simultaneamente ao mecanismo da resposta imune humoral, inicia-se a ativação da resposta imune celular, através da participação das células apresentadoras de antígenos (APCs). As APCs mais importantes nesse processo são os macrófagos e células dendríticas (DCs), que após o processamento do antígeno apresentam os peptídeos processados através do complexo MHC-II/peptídeo para os linfócitos T $CD4^+$. É importante ressaltar que a interação entre a APC e o linfócito T dependerá da ligação específica com o receptor de superfície de células T (TCR), da presença da molécula $CD4^+$, bem como de moléculas coativadoras. Essa interação precisa acontecer de forma precisa para que a resposta imune seja efetiva (PROIETTI, 2006).

Os monócitos podem ser classificados em subpopulações de acordo com o padrão de expressão das moléculas CD14 e CD16: clássicos ($CD14^{++}$, $CD16^-$), intermediários ($CD14^{++}$ $CD16^+$) e não clássicos ($CD14^+$, $CD16^{++}$). Apesar de não ter sido demonstrada diferença na frequência dessas subpopulações em pacientes não infectados e pacientes assintomáticos, pacientes com HAM/TSP possuem uma frequência menor de monócitos clássicos e maior de monócitos intermediários (AMORIM et al., 2014).

Já as moléculas indicadoras de ativação celular CD80, CD86 e HLA-DR, presentes na superfície dos monócitos, são expressas de maneira similar em pacientes com diagnóstico positivo e negativo para HAM/TSP, bem como em indivíduos não infectados pelo HTLV-1 (AMORIM et al., 2014). O mesmo não acontece com as células dendríticas, já que os pacientes infectados pelo HTLV-1 apresentam uma diminuição na expressão de CD14 e CD1a, além de moléculas como CD83, CD86 e HLA-DR, mesmo após estimulação com TNF (JAIN et al., 2009; NASCIMENTO et al., 2011). Estudos anteriores demonstram também que macrófagos de pacientes com HTLV-1 produzem menos IL-10, quando comparados com indivíduos não infectados (AMORIM et al., 2014).

Após a apresentação do antígeno pela APCs, as células T $CD4^+$ ativadas passam a produzir citocinas e, consequentemente, a ativar e potencializar as respostas imunológicas, incluindo a produção de anticorpos e a resposta citotóxica. Essa resposta citotóxica, realizada pelos linfócitos T $CD8^+$ ocorre a partir da liberação de granzima e perforina, capazes de lisar a célula infectada (COOK et al., 2013). A eficiência dessa lise celular influencia diretamente a carga proviral do hospedeiro e pode ser quantificada a partir da marcação da proteína CD107a, presente no interior da vesícula que armazena a granzima e perforina.

Sabe-se que a maioria dos linfócitos T $CD8^+$ específicos para o HTLV-1 reconhecem epítopos da proteína Tax, mas outros antígenos também exercem um papel importante no desenvolvimento da resposta imune (JACOBSON et al., 1990; KANNAGI et al., 1991). Vale ressaltar ainda que embora o HTLV-1 apresente um tropismo por células T $CD4^+$, ele é capaz de infectar células T $CD8^+$ (HANON et al., 2000), e outras células não-T, incluindo monócitos e linfócitos B (KOYANAGI et al., 1993), macrófagos (NATH, 2002), células dendríticas (KNIGHT et al., 1993) e endoteliais (SETOYAMA et al., 1998).

Pise-Masison e colaboradores avaliaram se a persistência viral pode ser influenciada por mutações na ORF-I do HTLV-1 capazes de alterar a concentração relativa das proteínas p12 e p8. A mutação N26 resulta em uma maior expressão de p8, enquanto a G29S aumenta a concentração da p12. Na figura 12, é possível observar a influencia das proteínas p12 e p8 na capacidade de lise dos linfócitos T $CD8^+$ (CTL), onde as células T $CD4^+$ com expressão equivalente de p12 e p8 (D26) apresentaram menores taxas de lise, enquanto as células com expressão alterada da ORF-I (N26 e G29S), bem como as células deficientes em p12 e p8 (12KO), foram eficientemente reconhecidas pelas CTLs (PISE-MASISON et al., 2014).

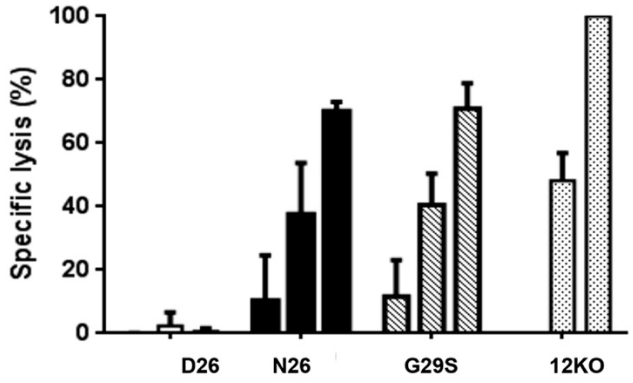


Figura 11. Taxa de lise específica de células T CD4+ infectadas com diferentes cepas do HTLV-1 PISE-MASISON et al., 2014.

Essas mutações também foram avaliadas quanto a sua influência na manutenção da carga proviral, onde foram observadas altas cargas provirais em pacientes com diagnóstico positivo e negativo para HAM/TSP que possuíam expressão equivalente de p12 e p8 (Figura 13). Esses dados sugerem que mutações na ORF-I podem influenciar a persistência viral (PISE-MASISON et al., 2014).

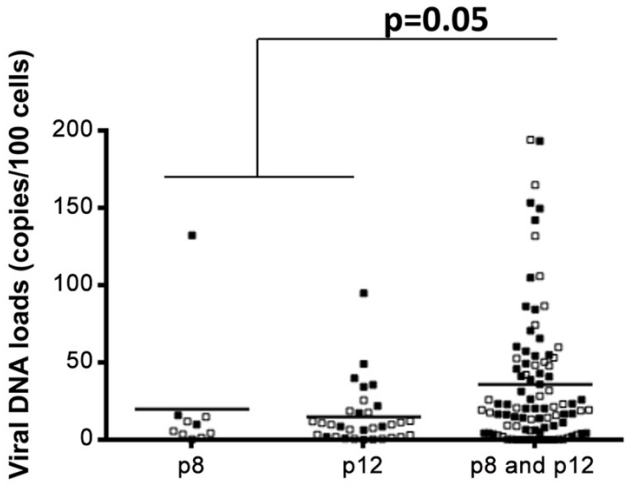


Figura 12. Análise de carga proviral em PBMC de pacientes infectados pelo HTLV-1, com mutações na ORF-1 Símbolos abertos = HC; símbolos fechados = HAM/TSP. PISE-MASISON et al., 2014.

3.8 VACINAS ANTI-HTLV-1

O desenvolvimento de vacinas tem proporcionado um grande impacto na saúde pública mundial. Entretanto, apesar da infecção pelo HTLV-1 ser endêmica em diferentes regiões geográficas do mundo, ainda permanece sem um método de profilaxia eficaz. A maioria das vacinas de DNA plasmidial, desenvolvidas até o momento, induziram resposta imune protetora em modelos utilizando animais de pequeno porte, porém a imunogenicidade destas em primatas mostrou-se menos potente (DONNELLY et al., 1997; MACGREGOR et al., 1998).

Estudos recentes sugerem uma nova abordagem, que otimiza elementos regulatórios no esqueleto do vetor vacinal, melhorando a imunogenicidade das vacinas de DNA. Entre diversas regiões analisadas, verificou-se que a adição de uma sequência reguladora da região R da região LTR do HTLV-1 à região promotora/reguladora do citomegalovírus (CMV) aumenta de 5 a 10 vezes a expressão dos transgenes *env* e *gag* do HIV-1, e intensifica a resposta imune celular contra antígenos. Foi demonstrado também que além da incorporação do elemento R do HTLV-1, a utilização dos transgenes separados, no lugar de genes codificando proteínas de fusão, representa uma estratégia simples e prática que contribui para o aperfeiçoamento das vacinas de DNA (BAROUCH et al., 2005). Estes dados demonstram que a otimização de elementos regulatórios específicos aumenta substancialmente a imunogenicidade das vacinas de DNA codificando antígenos diferentes em animais de pequeno porte e em primatas não humanos.

Em outro estudo, realizado por um grupo de pesquisadores do Institut Pasteur da França e do Instituto Nacional do Câncer nos EUA, foi avaliada a imunogenicidade e eficácia de uma vacina de DNA, cujo vetor era derivado do vírus vaccinia (NYVAC) contendo o gene *env* ou *env* e *gag* do HTLV-1. Três macacos-esquilo foram vacinados, com o recombinante contendo apenas *env* (doses nos tempos 0, 1 e 3 meses) e submetidos à transfusão intravenosa de células infectadas com o HTLV-1. A resposta imune contra o HTLV-1 foi detectada em um único animal vacinado. No entanto, após uma dose única (inicial) com o recombinante contendo apenas *env*, seguido pela imunização com a vacina NYVAC HTLV-1 *gag* e *env*, aos 6, 7 e 8 meses, todos os três animais se mostraram protegidos após transfusão com células infectadas pelo HTLV-1(KAZANJI et al., 2001).

Além disso, estudos avaliaram a eficácia de um vetor vacinal derivado do poxvírus (R-ALVAC), onde quatro coelhos brancos da Nova Zelândia foram vacinados com o vetor R-ALVAC nos meses 0 e 1. Após o desafio vacinal, 100% dos coelhos mostraram-se protegidos. Infelizmente, essa proteção não foi duradoura, visto que no mês 11 os animais se tornaram infectados (FRANCHINI et al., 1995).

Até o presente momento não foram descritas nenhuma profilaxia e terapia específicas eficazes na cura da infecção pelo HTLV-1. Apesar da alta morbidade, os pacientes que desenvolvem manifestações clínicas devem ser acompanhados e submetidos a metodologias paliativas, como tratamentos fisioterápicos e/ou medicamentosos.

4 RESULTADOS

Os resultados obtidos sobre a caracterização molecular da ORF-I do HTLV-1 proveniente de pacientes com ATLL, DIH, HAM/TSP e assintomáticos, bem como a anotação do principal genoma completo do HTLV-1 foram organizados em três artigos.

O primeiro artigo, intitulado "*A fully annotated genome sequence of human T-cell lymphotropic virus type 1 (HTLV-1)*", descreve a anotação completa do principal clone molecular do HTLV-1 (ATK1), utilizado como sequência referência nas análises *in silico* para a caracterização molecular da ORF-I. Neste artigo, publicado na revista "*Journal of Bioinformatics, Computational and Systems Biology*", foi descrita a posição nucleotídica inicial e final dos genes do HTLV-1 e seus respectivos produtos.

O segundo artigo, publicado na revista "*Infection, Genetics and Evolution*", contempla os resultados sobre a correlação entre mutações na ORF-I do HTLV-1, carga proviral e desenvolvimento de HAM/TSP, e possui como título "*Analyses of HTLV-1 sequences suggest interaction between ORF-I mutations and HAM/TSP out*come". Os resultados obtidos nesse trabalho sugerem uma potencial associação entre mutações específicas na ORF-I, alta carga proviral e desenvolvimento de HAM/TSP.

O terceiro artigo, intitulado "*Assessment of genetic diversity of HTLV-1 ORF-I sequences collected from patients with different clinical profiles*", aborda a avaliação da diversidade genética da ORF-I do HTLV-1 em pacientes com diferentes perfis clínicos e se encontra em processo de preparação para publicação.

4.1 A fully annotated genome sequence of human T-cell lymphotropic virus type 1 (HTLV-1).

Bioinformatics Computational& Systems Biology

http://elynsgroup.com

Short Communication Journal of Bioinformatics, Computational and Systems Biology Open Access

A Fully Annotated Genome Sequence of Human T-Cell Lymphotropic Virus Type 1 (HTLV-1)

Fernanda K. Barreto[1], Thessika HA. Araújo[1,2], Filipe F. Almeida Rego[3], and Luiz CJ. Alcantara

[1]*Oswaldo Cruz Foundation, Institute Gonçalo Moniz , Salvador, Bahia, Brazil*

[2]*Bahian School of Medicine and Public Health, Salvador, Bahia, Brazil*

[3]*Catholic University of Salvador, Salvador, Bahia, Brazil*

Received Date: May 09, 2017, **Accepted Date:** June 16, 2017, **Published Date:** June 23, 2017.

***Corresponding author:** Luiz Carlos Junior Alcantara, Oswaldo Cruz Foundation, Institute Gonçalo Moniz, Salvador, Bahia, Brazil, E-mail: lalcan@bahia.fiocruz.br .*

Abstract

In the next generation of genome sequencing, sequence annotation plays an important role with respect to genome evaluation. The aim of annotation is to identify key features in the genome, such as genes and their products. Although annotation tools are available and some sequence features have been published, annotation information for many complete and partial genomes of Human T-Cell Lymphotropic Virus Type 1 (HTLV-1) remains unavailable from GenBank. Sequence analysis is critical to the understanding of the pathogenesis of HTLV-1, and a well-annotated reference sequence is an essential component in this analysis. More accurate and complete information about the HTLV-1 genome can assist the scientific community in investigations on possible therapeutic and prophylactic vaccines, as well as aid studies on the pathogenesis of HTLV-1-associated diseases. Here we describe for the first time the complete nucleotide position annotation of the frequently used HTLV-1 reference sequence, ATK1 (accession number: J02029.1).

Keywords: HTLV-1; ATK1; Complete Genome; Annotation

Introduction

Human T-cell lymphotropic virus type 1 (HTLV-1) is present throughout the world and it is estimated that 5–10 million individuals are infected [1]. This retrovirus has been mainly linked to Adult T Cell Leukemia/Lymphoma (ATLL), Tropical Spastic Paraparesis/HTLV-associated myelopathy (HAM/TSP) and infective dermatitis [2–4].

One of the challenges faced by researchers in the development of an HTLV-1 vaccine is to determine why some individuals develop pathological processes, while others remain asymptomatic. Genomic studies have indicated that HTLV-1 mutations may be associated with infection outcome, yet the GenBank database contains relatively few complete genomes available [5–8]. In addition, the most used HTLV-1 genome (ATK1) is incomplete with regard to the start and end nucleotide position of each gene. ATK1 was the first human retrovirus genome described and to date has not been fully annotated [9]. Here, we performed the complete nucleotide position annotation of the full ATK1 genome available at GenBank. We hope that this information will support future HTLV-1 research efforts by the scientific community.

Materials and Methods

To perform the complete nucleotide position annotation of the most used HTLV-1 genome (ATK1), this sequence was downloaded from GenBank (accession number: J02029.1) and all available features were recorded. Next, we identified in GenBank other complete and partial HTLV-1 sequences with some nucleotide position information, through the "HTLV-1 complete sequence; HTLV-1 and LTR; HTLV-1 and HBZ; HTLV-1 and p12; HTLV-1 and p30" keywords.

After downloading these sequences, Clustal X 2.0 software was used to align all sequences, including ATK1 [10,11]. The alignment was manually edited and the correct nucleotide positions of the HTLV-1 genes in the complete and partial sequences was analyze in relation to ATK1 sequence. The nucleotide position annotation of ATK1 was performed using Geneious R6 software [12]. Finally, Universal Protein Resource (UniProt, www.uniprot.org) was used to confirm coding region annotations through the alignment of HTLV-1 protein sequences available in the UniProt and the ATK1 sequence translated based on our annotations [13]. Figure 1 explains the workflow of ATK1 nucleotide annotation.

Results and Discussion

Complete nucleotide annotations provide the scientific community with the necessary data to better interpret biological processes. In the case of HTLV-1, this information is particularly important since the literature is controversial with respect to the nucleotide position of each of the protein products, especially those produced by pX.

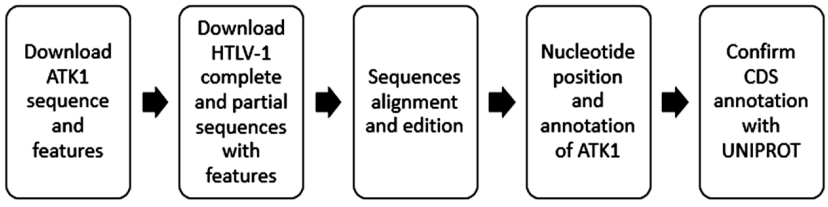


Figure 1: The workflow of ATK1 nucleotide position annotation. The ATK1 accession numbers is J02029.1. The accession numbers of the HTLV-1 complete and partial sequences downloaded are NC_001436, Y16487.2, U19949, JX184913.1 and KM436104.1.

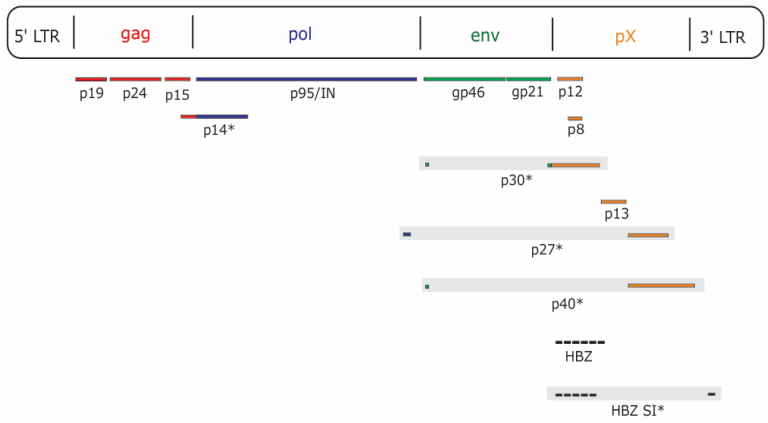


Figure 2: The fully annotated genome of human T-cell lymphotropic virus type 1 (HTLV-1). (The complete HTLV-1 genome is represented as a box. The LTR promoter regions and the genes encoded by sense mRNA are shown in different colors. Each protein is represented by a line with a color specific to its corresponding gene. HBZ and HBZ-SI are encoded by an antisense mRNA, represented by a dotted line. Proteins labeled with asterisks (*) are encoded by more than one gene).

Region	Repeat Features		Location
5' LTR	U3		23....374
	R		375...603
	U5		604...777
3' LTR	U3		8301...8652
	R		8653...8881
	U5		8882...9055
Gene	**Mature Peptides**	**CDS**	**Location**
gag	gag-pro-pol	p19	824...1213
		p24	1214...1855
		p15	1856...2113
pro	gag-pro-pol	p14	1960...2778
pol	gag-pro-pol	p95/IN	-...5210
env	gp63	gp46	5203...6138
		gp21	6139...6669
pX	p12	p12	6857...7156
	p8	p8	6944...7156
	p30	p30	5203-5206...6853-7574
	p13	p13	7311...7574
	Rex	p27	5147-5206...7325-7834
	Tax	p40	5203-5206...7325-8382
	HBZ	HBZ	7312...6686
	HBZ SI	HBZ SI	7290-6686...8702-8690

Table 1: Nucleotide positions of HTLV-1 genes in the ATK1 sequence (accession number: J02029.1). (The sites of reverse transcriptase (p95) and integrase translation initiation have not been determined; LTR = long terminal repeat).

To establish a localization standard for the HTLV-1 genes and proteins, we analyzed the ATK1 sequence and performed the complete nucleotide annotation of this sequence. As shown in figure 2, the HTLV-1 genome is composed of genes gag, pol, env and the pX region, flanked by two Long Terminal Repeat (LTR) regions at both 5' and 3' ends. The gag precursor protein is cleaved into products. The pol gene encodes polymerase p95 (reverse transcriptase) and an integrase, although the sites of translation initiation have yet to be determined. A frameshift occurs at the 3' gag termination and the beginning of the pol gene, which encodes p14 (protease). The env precursor protein is also cleaved to generate two products: gp46 and gp21. The pX region contains four overlapping open reading frames (ORF) that encode regulatory and accessory proteins and an antisense mRNA that generates the basic leucine zipper (HBZ) protein and the isoform of HBZ (HBZ-SI). ORF-I produces the p12 protein, which can be further cleaved into p8 protein, while ORF-II produces two proteins: p13 and p30, with part of the p30 protein being coded by env. ORF-III and ORF-IV produce proteins p27 (Rex) and p40 (Tax), respectively, both also partially coded by env. The size of this genome is approximately 9 kilobase (kb) and the start and end nucleotide positions of each gene are described below in Table 1.

The RefSeq database of GenBank suggests another sequence as reference (accession number: NC_001436). However, most of HTLV-1 papers used the ATK1 as reference sequence in their analysis [14–16]. Nevertheless, both of these sequences do not have information about all the HTLV-1 products, as p14, p12, p8, p30, p13 and HBZ. Therefore, our complete results can be used as a reference for the alignment and annotation of other HTLV-1 genomes.

Conclusion

The present study attempted to perform a complete nucleotide annotation of the most used HTLV-1 complete genome, ATK1. There are many questions that remain to be answered in the field of HTLV-1 research, and we hope that these data will assist other investigations carried out by the scientific community.

Availability of Data and Material

All sequences are available in the GenBank database (accession numbers: J02029.1; NC_001436; Y16487.2; U19949; JX184913.1; KM436104.1).

Authors' Contribution

All authors wrote, read and approved the final manuscript.

Conflicts of Interest

The authors declare that they have no competing interests.

References

1. Gessain A, Cassar O. Epidemiological Aspects and World Distribution of HTLV-1 Infection. Front Microbiol. 2012;3:388. doi: 10.3389/fmicb.2012.00388.
2. Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, et al. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet. 1985;2(8452):407-10.
3. Yoshida M, Miyoshi I, Hinuma Y. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc Natl Acad Sci U S A. 1982;79(6):2031-5.
4. Goncalves DU, Guedes ACM, Proietti AB de FC, Martins ML, Proietti FA, Lambertucci JR. Dermatologic lesions in asymptomatic blood donors seropositive for human T cell lymphotropic virus type-1. Am. J. Trop. Med. Hyg. 2003;68:562–565.
5. Ilinskaya A, Heidecker G, Derse D. Opposing effects of a tyrosine-based sorting motif and a PDZ-binding motif regulate human T-lymphotropic virus type 1 envelope trafficking. J Virol. 2010;84(14):6995-7004. doi: 10.1128/JVI.01853-09.
6. Barreto FK, Khouri R, Rego FF, Santos LA, Castro-Amarante MF, Bialuk I, et al. Analyses of HTLV-1 sequences suggest interaction between ORF-I mutations and HAM/TSP outcome. Infect Genet Evol. 2016;45:420-425. doi: 10.1016/j.meegid.2016.08.020.
7. Neto WK, Da-Costa AC, de Oliveira AS, Martinez VP, Nukui Y, Sabino EC, et al. Correlation between LTR point mutations and proviral load levels among Human T cell Lymphotropic Virus type 1 (HTLV-1) asymptomatic carriers. Virol. J. 2011;8:535. doi: 10.1186/1743-422X-8-535.
8. Magri MC, Costa EAS, Caterino-de-Araujo A. LTR point mutations in the Tax-responsive elements of HTLV-1 isolates from HIV/HTLV-1-coinfected patients. Virol J. 2012;9:184. doi: 10.1186/1743-422X-9-184.
9. Seiki M, Hattori S, Hirayama Y, Yoshida M. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA. 1983;80:3618–3622.
10. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8.
11. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2013;41:D36–42. doi: 10.1093/nar/gks1195.
12. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-9. doi: 10.1093/bioinformatics/bts199.
13. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge A, et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. In: Edwards D, editor. Plant Bioinforma. [Internet]. Springer New York; 2016 [cited 2017 Jan 9]. p. 23–54. Available from: https://link.springer.com/protocol/10.1007%2F978-1-4939-3167-5_2.
14. Pessôa R, Watanabe JT, Nukui Y, Pereira J, Casseb J, de Oliveira AC, et al. Molecular characterization of human T-cell lymphotropic virus type 1 full and partial genomes by Illumina massively parallel sequencing technology. PLoS One. 2014;9(3):e93374. doi: 10.1371/journal.pone.0093374.
15. Rosadas C, Vicente ACP, Zanella L, Cabral-Castro MJ, Peralta JM, Puccioni-Sohler M. Evidence of Sexual Transmission of Human T Cell Lymphotropic Virus Type 1 with ORF-I G29S Mutation Among Humans. AIDS Res Hum Retroviruses. 2017;33(4):328-329. doi: 10.1089/AID.2016.0238.
16. Mota-Miranda AC, de-Oliveira T, Moreau DR, Bomfim C, Galvão-Castro B, Alcantara LC. Mapping the molecular characteristics of Brazilian human T-cell lymphotropic virus type 1 Env (gp46) and Pol amino acid sequences for vaccine design. Mem. Inst. Oswaldo Cruz. 2007;102:741–9.

Corresponding author: Luiz Carlos Junior Alcantara, Oswaldo Cruz Foundation, Institute Gonçalo Moniz, Salvador, Bahia, Brazil, E-mail: lalcan@bahia.fiocruz.br.

Received Date: May 09, 2017, **Accepted Date:** June 16, 2017, **Published Date:** June 23, 2017.

Citation: Barreto FK, Araújo THA, Rego FFA, Alcantara LC (2017) A Fully Annotated Genome Sequence of Human T-Cell Lymphotropic Virus Type 1 (HTLV-1). J Bioinf Com Sys Bio 1(1): 105.

4.2 Analyses of HTLV-1 sequences suggest interaction between ORF-I mutations and HAM/TSP outcome.

Infection, Genetics and Evolution 45 (2016) 420–425

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Analyses of HTLV-1 sequences suggest interaction between ORF-I mutations and HAM/TSP outcome

Fernanda Khouri Barreto [a,1], Ricardo Khouri [b,1], Filipe Ferreira de Almeida Rego [a], Luciane Amorim Santos [a], Maria Fernanda de Castro-Amarante [c], Izabela Bialuk [d], Cynthia A. Pise-Masison [c], Bernardo Galvão-Castro [a,e], Antoine Gessain [f], Steven Jacobson [g], Genoveffa Franchini [c], Luiz Carlos Alcantara Jr [a,*]

[a] Centro de Pesquisa Gonçalo Moniz-Fiocruz, Laboratório de Hematologia Genética e Biologia Computacional, Salvador, Bahia, Brazil
[b] KU Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory of Clinical and Epidemiological Virology, Leuven, Belgium
[c] Animal Models and Retroviral Vaccines Section, National Cancer Institute, Bethesda, USA
[d] Department of General and Experimental Pathology, Medical University in Białystok, Poland
[e] Escola Bahiana de Medicina e Saúde Publica, Salvador, Bahia, Brazil
[f] Unité d'Epidémiologie et Physiopathologie des Virus Oncogènes, Département de Virologie, Batiment Lwoff, Institut Pasteur, Paris, France
[g] Viral Immunology Section, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, Bethesda, USA

ARTICLE INFO

Article history:
Received 5 April 2016
Received in revised form 17 August 2016
Accepted 19 August 2016
Available online 21 August 2016

Keywords:
HTLV-1
ORF-1
HAM/TSP

ABSTRACT

The region known as pX in the 3′ end of the human T-cell lymphotropic virus type 1 (HTLV-1) genome contains four overlapping open reading frames (ORF) that encode regulatory proteins. HTLV-1 ORF-I produces the protein p12 and its cleavage product p8. The functions of these proteins have been linked to immune evasion and viral infectivity and persistence. It is known that the HTLV-1 infection does not necessarily imply the development of pathological processes and here we evaluated whether natural mutations in HTLV-1 ORF-I can influence the proviral load and clinical manifestation of HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP). For that, we performed molecular characterization, datamining and phylogenetic analysis with HTLV-1 ORF-I sequences from 156 patients with negative or positive diagnosis for HAM/TSP. Our analyses demonstrated that some mutations may be associated with the outcome of HAM/TSP (C39R, L40F, P45L, S69G and R88K) or with proviral load (P34L and F61L). We further examined the presence of mutations in motifs of HBZ and observed that P45L mutation is located within the HBZ nuclear localization signal and was found more frequently between patients with HAM/TSP and high proviral load. These results indicate that some natural mutations are located in functional domains of ORF-I and suggests a potential association between these mutations and the proviral loads and development of HAM/TSP. Therefore it is necessary to conduct functional studies aimed at evaluating the impact of these mutations on the virus persistence and immune evasion.

1. Introduction

The human T-cell lymphotropic virus type 1 (HTLV-1) was the first human retrovirus identified that is associated with disease, including adult T cell leukemia/lymphoma (ATLL), tropical spastic paraparesis/HTLV-associated myelopathy (HAM/TSP), and infective dermatitis, polymyositis (Gessain et al., 1985; Goncalves et al., 2003; Poiesz et al., 1980; Yoshida et al., 1982). A recent research based on individuals originated from known HTLV-1 endemic areas estimates 5–10 million infected individuals (Gessain and Cassar, 2012). Indeed, HTLV-1 infection does not necessarily imply the development of pathological processes and the majority of infected individuals remain asymptomatic, with only a low percentage of individuals developing HAM/TSP (0.3–2%) (Kaplan et al., 1990; Maloney et al., 1998). Studies indicate that high proviral load is typically seen in individuals with a positive diagnosis of HAM/TSP, when compared with infected individuals without the diagnosis for this disease (Nagai et al., 1998).

A region near the 3′ end known as pX encodes important regulatory and accessory proteins shown to effect viral replication mechanisms (Edwards et al., 2011; Lairmore et al., 2012, 2011). Transcription of this region results in alternative forms of messenger RNA (mRNA), which contains four partially overlapping open reading frames (ORFs) and an antisense mRNA that generates the basic leucine zipper (HBZ) protein (Berneman et al., 1992; Ciminale et al., 1992; Gaudray et al., 2002; Koralnik et al., 1992). Interestingly the promoter of the four ORFs is 5′ LTR, but the HBZ promoter is in 3′ LTR (Berneman et al., 1992). Singly

* Corresponding author.
E-mail address: lalcan@bahia.fiocruz.br (L.C. Alcantara).
[1] Equal contributors.

http://dx.doi.org/10.1016/j.meegid.2016.08.020
1567-1348/

spliced mRNA from ORF-I produces the protein p12 that can be further cleaved into the p8 protein. Differential splicing of mRNA from ORF-II encodes the proteins, p30 and p13. The ORF-III and ORF-IV produce the proteins Rex and Tax, respectively (Ciminale et al., 1992; Johnson et al., 2001; Kiyokawa et al., 1985, 1984; Koralnik et al., 1992).

The expression of ORF-I in infected cells has been shown to regulate the biologically activity of HTLV-1, influencing the virus infectivity and persistence (Pise-Masison et al., 2014). The p12 protein resides in the endoplasmic reticulum and its expression increases the intracellular calcium concentration from the ER through inositol trisphosphate receptors, consequently promoting the dephosphorylation of activate nuclear factor of activated T-cells (NFAT) (Albrecht et al., 2002; Ding et al., 2002; Fukumoto et al., 2009). In addition, p12 can bind newly synthesized major histocompatibility complex class I (MHC I) molecules, reducing its expression on the cell surface. Although downregulation of MHC class I triggers NK cell activation, HTLV-1 infected cells are resistant to NK cell killing through p12's reduction of cell surface ICAM-1 and ICAM-2 (Banerjee et al., 2007).

The p12 protein undergoes double proteolytic cleavage. The first cleavage occurs between amino acid positions 9 and 10 and removes the ER retention/retrieval signal of p12. The second cleavage occurs between amino acids 29 and 30 and generates the p8 protein (Edwards et al., 2011). The p8 protein resides at the cell surface and its expression induces cellular conduits and enhances virus transmission (Van Prooyen et al., 2010).

Trovato et al. demonstrated two natural allelic variants of p12: a lysine-to-arginine change at position 88, which was found more frequently in patients with HAM/TSP and could be important for HTLV-1 outcome (Trovato et al., 1999). More recent studies on ORF-I suggest the co-dependence of p12 and p8 proteins to virus persistence and demonstrated that natural mutations within ORF-I sequences can affect the relative amounts of these proteins (Pise-Masison et al., 2014). Here, we demonstrate that natural mutations in HTLV-1 ORF-I might influence the proviral load and clinical manifestation of HAM/TSP.

2. Materials and methods

To perform the molecular analysis of ORF-I HTLV-1 we analysed 1530 HTLV-1 ORF-I sequences of different isolated cell clones from a total of 156 HTLV-1-infected subjects. These samples were obtained from the Centre Hospitalier Universitaire de Fort-de-France in Martinique, the Institut Pasteur de Cayenne in French Guyana, the Bahia School of Medicine and Public Health and the National Institutes of Health Clinical Center. Blood samples from these individuals were obtained and the DNA extracted from PBMCs was used to determine the proviral load using Real-time PCR of Tax. The same DNA was used for amplification of ORF-I through PCR reactions and these PCR products was purified and cloned into specific vector. Five to twenty clones per patient were isolated and sequenced. These methods were described in detail by Pise-Masison et al.(Pise-Masison et al., 2014).

The final dataset was composed of 879 HTLV-1 ORF-I sequences from 86 patients with not HAM/TSP, referred to as Health Carrier (HC), and 651 HTLV-1 ORF-I sequences from 70 patients with HAM/TSP. These sequences were originated from different geographic regions (Caribbean, France, North America, Africa, and Brazil) and classified according to a negative or positive diagnostic for HAM/TSP. All samples were anonymized and the clinical classification was carried out by medical experts according to World Health Organization (WHO). The research was conformed to the guidelines of the ethics review board of the National Cancer Institute and informed consent was written and obtained from each subject.

2.1. Molecular characterization

The genetic distances among the sequences from patient with negative or positive diagnosis for HAM/TSP were measured within and between the different datasets using the MEGA 5.05 program and the analyses about the proviral load was conducted using the STATA program. The statistical analyses of the proviral load were performed using Mann-Whitney test and a P value lower than 0.05 was considered statistically significant.

Since we get differing numbers of clones from each patient, the characterization of mutations was determined by a qualitative analysis considering the presence or absence of variants among the sequences of each individual. If one of the patient sequences has the variant we consider that the patient has the mutation (Geneious 5.6.5 program). We selected the most frequent natural mutations within ORF-I sequences (F3L, S23P, D26N, G29S, P34L, C39R, L40F, P45L, F61L, S63P, L66P, S69G, R83C, R88K and P91S) to evaluate their possible association with the development of HAM/TSP using the Excel program. The statistical analyses were performed using Fisher's exact test and a P value lower than 0.05 was considered statistically significant.

2.2. Datamining

We used a two-step approach, using an unsupervised method for attribute selection (naïve Bayesian networks analysis), followed by a supervised method for disease progression classification (Decision tree J48). A Bayesian network (BN) is a probabilistic model that describes statistical conditional dependencies between multiple variables. In this study, we learn Bayesian networks from observation of the attributes. Dependencies are visualized in a directed acyclic graph and form the qualitative component of the BN. In this graph, each node corresponds to an attribute, and a direct arc between nodes represents a direct influence. Mathematically, a Bayesian network provides a refactoring of the Joint Probability Distribution (JPD) of the data, using Bayes' rules. As a BN simplifies the JPD, it provides an effective model that summarizes statistical properties of the data. In this way, the best Bayesian network is searched that explains a maximum of the observed associations in the data using a minimum number of direct influences. Bayesian network learning was performed using the B-course software adapted by Deforche et al. (Deforche et al., 2006). In this non-linear model, the conditional dependency was assessed with a nonparametric bootstrap (100× replicates) (Friedman et al., 2013). The Decision tree J48 strategy is based on divide-and-conquer approach, sometimes called top-down induction of decision trees. It was developed and refined by J. Ross Quinlan at the University of Sydney in Australia. Divide-and-conquer algorithms operate by adding tests to the tree that is under construction, always striving to maximize the separation between the classes. Each of these involves finding an attribute to split on. The state of the art for error evaluation, independently from the loss functions, is to execute the 10-fold cross validation when a test set is unavailable. To execute and validate J48 decision tree approach we used the data mining suite Weka (E. WiaF, 2011).

2.3. Phylogenetic analysis

The phylogenetic analysis was performed using consensus sequences of patients from Brazil. Clustal-X was used for the alignment and the TN93 model of nucleotide substitution was selected using jModelTest. The maximum likelihood tree was inferred using PhyML online tool and bootstrap analysis (1000 replicates) was used to calculate the statistical support of the tree branches. Tree visualization and editing was done using FigTree v.1.2.2.

3. Results

3.1. Descriptive characteristics of the study samples

In this study, we analysed HTLV-1 ORF-I sequences from a total of 156 HTLV-1-infected subjects. Our final dataset was composed of 879

HTLV-1 ORF-I sequences from independently isolated clones of 86 non-HAM/TSP individuals (HC) and 651 HTLV-1 ORF-I sequences from independently isolated clones of 70 patients diagnosed with HAM/TSP (Pise-Masison et al., 2014). The median proviral load of patients with HAM/TSP was 22.9 copies/100 cells (IQR 11.9–47.6) which was significantly higher than HC, whose median proviral load was 8.7 copies/100 cells (IQR 0,5–17,8) ($p < 0,0001$).

Next, we measured the genetic distance between sequences using the MEGA 5.05 program. Interestingly, we found that the diversity between sequences from patients with HAM/TSP (0.014) was lower than that found between sequences from heath carrier (0.017).

3.2. Association of mutations in ORF-I sequences from HTLV-1-infected patients with proviral load and clinical definition

We analysed the presence or absence of fifteen amino acid changes (F3L, S23P, D26N, G29S, P34L, C39R, L40F, P45L, F61L, S63P, L66P, S69G, R83C, R88K and P91S) that influence the expression profile of the HTLV-1 ORF-I protein product. As previously described (Pise-Masison et al., 2014), these mutations resulted in three patterns of ORF-I protein expression: higher levels of p8 expression, higher levels of p12 expression or an equivalent level of p8 and p12 expression. We then compared the frequency of each mutation in non-HAM/TSP

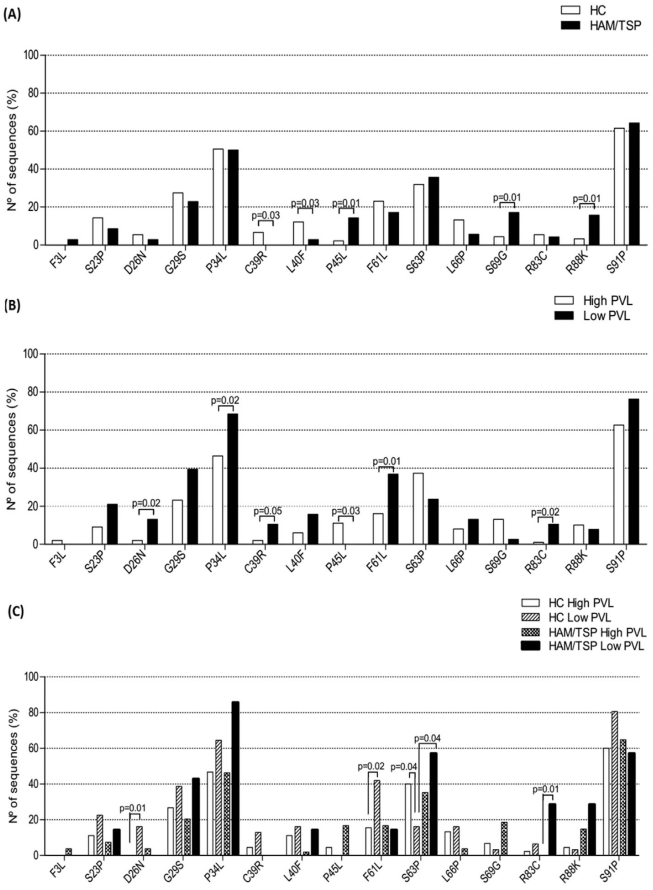


Fig. 1. Frequency of natural mutations within ORF-I sequences. The characterization of mutations was determined by a qualitative analysis considering the presence or absence of variants among the clone sequences of each individual. (A) The graph represents a comparison between patients with positive or negative diagnosis of HAM/TSP for a given mutation. (B) Comparison between patients with high and low proviral loads (PVL). High PVL are indicated by white bars and the patients with low PVL are indicated by black bars. (C) Mutation frequency was graphed for HAM/TSP patients with high PVL (black bar), HAM/TSP with low PVL (dotted bar), no HAM/TSP (HC) with high PVL (white bar) and HC with low PVL (striped bar). The phenotype of ORF-I protein expression is: Mostly expression of p12 (F3L, S23P, G29S, P34L, C39R,L40F, P45L, F61L, S63P, L66P); Mostly expression of p8 (D26N); Same expression of p12 and p8 (S69G, R83C, R88K, S91P). P values were calculated by the Fisher's exact test.

patients (HC) to HAM/TSP patients. In addition, a comparison between amino acid sequence and proviral load was performed (Fig. 1).

As shown in Fig. 1A, mutations P45L, S69G and R88K were found more frequently in patients with a positive diagnosis for HAM/TSP, whereas mutations C39R and L40F were found more frequently in HC. Of note, Trovato et al. showed that a lysine residue at position 88 destabilized the p12 protein by proteasomal degradation and that this mutation could be associated with HAM/TSP (Trovato et al., 1999). While this mutation was found more frequently in HAM/TSP patients than HC, we did not find a correlation between R88K and proviral load (Fig. 1B).

In contrast, mutations D26N, P34L, C39R, F61L and R83C were associated with low proviral load and P45L with high proviral load (Fig. 1B). A combined classification of proviral load and clinical diagnosis reveals mutations D26N and F61L are associated with low proviral load and HC, whereas mutations S63P and R83C are associated with high proviral load and HAM/TSP (Fig. 1C). The mutations P34L and S91P appear most frequently (>than 40% of the sequences analysed) but are not associated with the development of HAM/TSP. However, as seen in Fig. 1B, there was a higher frequency of P34L sequences in individuals with low proviral load.

Similarly, mutations C39R and F61L were more frequently found in patients with low proviral load (Fig. 1B). These three mutations (P34L, C39R and F61L) that result in predominant expression of p12 are associated with low proviral load. Whereas, the mutation D26N, the only mutation that resulted in predominant p8 expression, was not associated with the development of HAM/TSP but was associated with low proviral load and HC (Fig. 1B and C). These results are consistent with that of Pise-Masison et al. (Pise-Masison et al., 2014), where predominant expression of either p12 or p8 was associated with low proviral load.

Low proviral load in some individuals might also be influenced by alterations in remaining open reading frames in which other accessory proteins are encoded, such as viral protein HTLV-1 bZIP factor (HBZ). As described in previous studies, HBZ is encoded by the complementary strand of the HTLV-1 genome (7292 to 6666) and seems to play a role in cellular transformation, survival and proliferation (GenBank: U19949) (Matsuoka, 2010; Matsuoka and Green, 2009). Here, we further examined the presence of these mutations in motifs of HBZ and only four mutations in ORF-I resulted in changes in the HBZ ORF: P34L, P45L, L66P and R88K. Among them, only the P45L mutation is located within the motif previously described (HBZ nuclear localization signal) and was found more frequently between patients with HAM/TSP and high proviral load.

3.3. Data mining approach reveals that mutation interactions are necessary to define HAM/TSP pathogenesis

To reveal possible interactions among specific mutations in ORF-I, proviral load, and disease definition, we employed a two-step approach, using an unsupervised method for attribute selection (naive Bayesian networks analysis, BN), followed by a supervised method for disease progression classification (Decision tree J48). First, Bayesian network analysis with all sequences showed a central association among disease progression, proviral load and patient origin, that interacted directly with mutations in ORF-I region: D26N, G29S, P34L, L40F, P45L, S63P, L66P, S69G, R83C. This first BN analysis revealed a possible biased geographical distribution of mutation prevalence given that in our dataset we do not have sequences from patients HC and HAM/TSP from all the geographic regions (data not shown). Due to this, we decided to use only samples from Brazil (27 HAM/TSP patients [median of 9 samples/patient, IQR = 7–17] and 48 HC [median of 10.5 samples/patient, IQR = 8–19]). No statistical significance was observed between the medians of the number of samples per HAM/TSP patients and HC (Fig. 2A). We then applied Decision tree J48 to predict HAM/TSP samples, confirming univariate analysis and revealing hidden conditional dependencies among mutations and proviral load (Fig. 2B). We observed several interactions associated with a true positive rate of 0.63 for HAM/TSP, as among D26N mutation and high proviral load or interactions between high proviral load, 26D, 29G and 34P. The predicted results above were plotted in a matrix 2 × 2 and Sensitivity, Specificity, Positive Predictive Values and Negative Predictive Values were described in Table 1.

Since sequences "states" are not statistically independent observations and these findings between mutations and disease manifestation could be associated with shared ancestry inherent, we performed a phylogenetic analysis using sequences of patients from Brazil (n = 75 consensus sequences) but the phylogenetic tree presented poor phylogenetic signal with no bootstrap support. This result might be explained due to the HTLV-1 ORF-I being a short region (300 bp) and highly conserved and it is known that the indicated region for subtyping HTLV-1 is LTR (Alcantara et al., 2009).

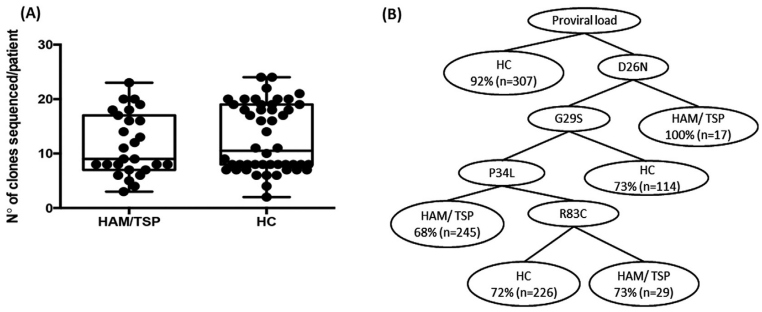


Fig. 2. Analysis of interactions among mutations in ORF-I, proviral load, and development of HAM/TSP. (A) Comparison of number of samples sequenced between HAM/TSP patients and HC from Brazil. This dataset is composed of 27 HAM/TSP patients [median of 9 samples/patient, IQR = 7–17] and 48 HC [median of 10.5 samples/patient, IQR = 8–19] (Mann Whitney test, p = 0.43). (B) Discovery of hidden conditional dependencies between proviral load and ORF-1 mutations associated with HAM/TSP. To reduce noise in the statistical analysis, proviral load measurements were discretized in 25% quartile and the results were represented as binary decision tree graphs. Nodes outline the chosen attributes and the number (n =) of instances and the accuracy defined per decision branch (%). This dataset is composed of 938 samples (Overall Accuracy = 78.25%, Detailed accuracy per HAM/TSP: True positive rate = 0.60, f-measure = 0.65 and AUC = 0.8).

Table 1
Results of conditional dependencies between proviral load, ORF-1 mutations and clinical status. The ORF-I sequences were originated from different isolated cell clones of HTLV-1-infected patients from Brazil. PPV means positive predictive value and NPV means negative predictive value.

		Sequences		Parameter (%)			
Conditional dependency: Proviral load (PVL) and mutation (AA)	Prediction	HC (n = 616)	HAM/TSP (n = 322)	Sensitivity	Specificity	PPV	NPV
Low PVL	HC	283	24	86	63	82	70
High PVL + 26N	HAM/TSP	0	17				
High PVL + 26D + 29S	HC	83	31				
High PVL + 26D + 29G + 34P	HAM/TSP	79	166				
High PVL + 26D + 29G + 34L + 83C	HAM/TSP	8	21				
High PVL + 26D + 29G + 34L + 83R	HC	163	63				

4. Discussion

HTLV-1 infection does not necessarily imply the development of pathological processes and it is not known what determines the manifestation of the disease in an infected individual. To date, proviral DNA levels in the blood are the best predictor of risk for the development of HAM/TSP (Matsuzaki et al., 2001; Yamano et al., 2002). Here, we described HTLV-1 ORF-I mutations that are associated with disease phenotype and may influence the outcome of infection and the development of HAM/TSP.

Computational analysis of the amino acid sequence of p12 predicts the existence of several functional domains, including a noncanonical endoplasmic reticulum (ER) retention signal, two leucine zipper (LZ) motifs, two transmembrane domains, a calcineurin-binding motif and four proline-rich Src homology 3 (SH3)-binding domains (Ding et al., 2002; Fukumoto et al., 2009). Some mutations analysed here are located in functional domains of ORF-I and may be associated with clonal expansion of HTLV-1-infected cells. The D26N and F61L mutations are located in the first and second transmembrane domains, respectively, and are found more frequently in patients with low proviral load, as shown in Fig. 1B. Our data support studies that suggest co-expression of p12 and p8 are important for efficient viral persistence (Pise-Masison et al., 2014).

Only R83C mutation is located in a calcineurin-binding motif and is associated with low proviral load. Calcineurin plays a crucial role in T-cell activation, in part, by dephosphorylating the nuclear factors of activated T cells (NFATs). NFAT is essential for activating cytokine gene expression and, thus, the immune response (Kim et al., 2003). p12 is able to mediate an increase in cytosolic calcium in T-cells by increasing calcium release from the ER (Kim et al., 2003). By depleting ER calcium stores and increasing cytosolic calcium, p12 can modulate cellular processes including T-cell proliferation, viral replication, and viral spread. Early studies on ORF-I showed that it activates NFAT (Albrecht and Lairmore, 2002; Kim et al., 2003). Therefore, mutations in the calcineurin binding region of ORF-I could affect NFAT activity and impact the immune response and viral replication.

Low proviral load might also be influenced by alterations in other accessory proteins, such as HBZ protein, that seems to play a role in cellular transformation, survival and proliferation (Albrecht and Lairmore, 2002; Gaudray et al., 2002; Matsuoka and Green, 2009). Here, we demonstrated only the P45L mutation is located within the HBZ nuclear localization signal and was found more frequently between patients with HAM/TSP and high proviral load. This finding is consistent with results showing that HBZ down-regulates Tax-induced HTLV-1 transcription and thus viral replication (Matsuoka, 2010; Matsuoka and Green, 2009).

Many different machine-learning tools have been used to predict therapy failure and conditional dependency among drug resistance mutations (Pineda-Pena et al., 2014; Prosperi and De Luca, 2012). Here, our data mining approach confirmed the association among HAM/TSP, proviral load and specific single mutations found with univariate analysis. Moreover, it revealed unexpected conditional dependencies among proviral load and mutations in ORF-1 associated with HAM/TSP. Despite the high prevalence of P34L mutation and its association with proviral load, this mutation alone could not determine disease. Only the data mining approach revealed the coordinate dependency between proviral load, P34L and P45L to be associated with HAM/TSP.

5. Conclusions

The results presented in this study suggest that some ORF-I natural mutations may be associated to HAM/TSP development and to the proviral load of HTLV-1-infected individuals. Is important to note that HTLV-1 have four open reading frames (ORFs) in the pX region and all regulatory proteins play a key role in viral pathogenesis. Therefore to investigate better the contribution of mutations found in HTLV-1 ORF-I to the development of HAM/TSP, it is necessary to conduct functional studies aimed at evaluating the impact of these mutations on the virus persistence and immune evasion.

Availability of supporting data

All sequences are available from the NCBI database (accession numbers KM436104-KM437632).

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by the Intramural Program at the National Cancer Institute, National Institutes of Health, Bethesda, MD.

References

Albrecht, B., Lairmore, M.D., 2002. Critical role of human T-lymphotropic virus type 1 accessory proteins in viral replication and pathogenesis. Microbiol. Mol. Biol. Rev. MMBR 66, 396–406 (table of contents).

Albrecht, B., D'Souza, C.D., Ding, W., Tridandapani, S., Coggeshall, K.M., Lairmore, M.D., 2002. Activation of nuclear factor of activated T cells by human T-lymphotropic virus type 1 accessory protein p12(I). J. Virol. 76, 3493–3501.

Alcantara, L.C.J., Cassol, S., Libin, P., Deforche, K., Pybus, O.G., Van Ranst, M., Galvao-Castro, B., Vandamme, A.-M., de Oliveira, T., 2009. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucleic Acids Res. 37, W634–W642. http://dx.doi.org/10.1093/nar/gkp455.

Banerjee, P., Feuer, G., Barker, E., 2007. Human T-cell leukemia virus type 1 (HTLV-1) p12I down-modulates ICAM-1 and −2 and reduces adherence of natural killer cells, thereby protecting HTLV-1-infected primary CD4+ T cells from autologous natural killer cell-mediated cytotoxicity despite the reduction of major histocompatibility complex class I molecules on infected cells. J. Virol. 81, 9707–9717. http://dx.doi.org/10.1128/JVI.00887-07.

Berneman, Z.N., Gartenhaus, R.B., Reitz, M.S., Blattner, W.A., Manns, A., Hanchard, B., Ikehara, O., Gallo, R.C., Klotman, M.E., 1992. Expression of alternatively spliced human T-lymphotropic virus type I pX mRNA in infected cell lines and in primary uncultured cells from patients with adult T-cell leukemia/lymphoma and healthy carriers. Proc. Natl. Acad. Sci. U. S. A. 89, 3005–3009.

Ciminale, V., Pavlakis, G.N., Derse, D., Cunningham, C.P., Felber, B.K., 1992. Complex splicing in the human T-cell leukemia virus (HTLV) family of retroviruses: novel mRNAs and proteins produced by HTLV type I. J. Virol. 66, 1737–1745.

Deforche, K., Silander, T., Camacho, R., Grossman, Z., Soares, M.A., Van Laethem, K., Kantor, R., Moreau, Y., Vandamme, A.-M., 2006. Analysis of HIV-1 pol sequences using

Bayesian Networks: implications for drug resistance. Bioinforma. Oxf. Engl. 22, 2975–2979. http://dx.doi.org/10.1093/bioinformatics/btl508.
Ding, W., Albrecht, B., Kelley, R.E., Muthusamy, N., Kim, S.-J., Altschuld, R.A., Lairmore, M.D., 2002. Human T-cell lymphotropic virus type 1 p12(I) expression increases cytoplasmic calcium to enhance the activation of nuclear factor of activated T cells. J. Virol. 76, 10374–10382.
Edwards, D., Fenizia, C., Gold, H., de Castro-Amarante, M.F., Buchmann, C., Pise-Masison, C.A., Franchini, G., 2011. Orf-I and Orf-II-encoded proteins in HTLV-1 infection and persistence. Virologie 3, 861–885. http://dx.doi.org/10.3390/v3060861.
Friedman, N., Goldszmidt, M., Wyner, A., 2013. Data Analysis with Bayesian Networks: A Bootstrap Approach (ArXiv13016695 Cs Stat).
Fukumoto, R., Andresen, V., Bialuk, I., Cecchinato, V., Walser, J.-C., Valeri, V.W., Nauroth, J.M., Gessain, A., Nicot, C., Franchini, G., 2009. In vivo genetic mutations define predominant functions of the human T-cell leukemia/lymphoma virus p12I protein. Blood 113, 3726–3734. http://dx.doi.org/10.1182/blood-2008-04-146928.
Gaudray, G., Gachon, F., Basbous, J., Biard-Piechaczyk, M., Devaux, C., Mesnard, J.-M., 2002. The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. J. Virol. 76, 12813–12822.
Gessain, A., Cassar, O., 2012. Epidemiological aspects and world distribution of HTLV-1 infection. Front. Microbiol. 3, 388. http://dx.doi.org/10.3389/fmicb.2012.00388.
Gessain, A., Barin, F., Vernant, J.C., Gout, O., Maurs, L., Calender, A., de Thé, G., 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet Lond. Engl. 2, 407–410.
Goncalves, D.U., Guedes, A.C.M., Proietti, A.B., de F.C., Martins, M.L., Proietti, F.A., Lambertucci, J.R., 2003. Dermatologic lesions in asymptomatic blood donors seropositive for human T cell lymphotropic virus type-1. Am. J. Trop. Med. Hyg. 68, 562–565.
Johnson, J.M., Harrod, R., Franchini, G., 2001. Molecular biology and pathogenesis of the human T-cell leukaemia/lymphotropic virus Type-1 (HTLV-1). Int. J. Exp. Pathol. 82, 135–147.
Kaplan, J.E., Osame, M., Kubota, H., Igata, A., Nishitani, H., Maeda, Y., Khabbaz, R.F., Janssen, R.S., 1990. The risk of development of HTLV-I-associated myelopathy/tropical spastic paraparesis among persons infected with HTLV-I. J. Acquir. Immune Defic. Syndr. 3, 1096–1101.
Kim, S., Ding, W., Albrecht, B., Green, P.L., Lairmore, M.D., 2003. A conserved calcineurin-binding motif in human T lymphotropic virus type 1 p12I functions to modulate nuclear factor of activated T cell activation. J. Biol. Chem. 278, 15550–15557. http://dx.doi.org/10.1074/jbc.M210210200.
Kiyokawa, T., Seiki, M., Imagawa, K., Shimizu, F., Yoshida, M., 1984. Identification of a protein (p40x) encoded by a unique sequence pX of human. Gan 75, 747–751.
Kiyokawa, T., Seiki, M., Iwashita, S., Imagawa, K., Shimizu, F., Yoshida, M., 1985. p27x-III and p21x-III, proteins encoded by the pX sequence of human T-cell leukemia virus type I. Proc. Natl. Acad. Sci. U. S. A. 82, 8359–8363.
Koralnik, I.J., Gessain, A., Klotman, M.E., Lo Monico, A., Berneman, Z.N., Franchini, G., 1992. Protein isoforms encoded by the pX region of human T-cell leukemia/lymphotropic virus type I. Proc. Natl. Acad. Sci. U. S. A. 89, 8813–8817.
Lairmore, M.D., Anupam, R., Bowden, N., Haines, R., Haynes 2nd, R.A.H., Ratner, L., Green, P.L., 2011. Molecular determinants of human T-lymphotropic virus type 1 transmission and spread. Virologie 3, 1131–1165. http://dx.doi.org/10.3390/v3071131.
Lairmore, M.D., Haines, R., Anupam, R., 2012. Mechanisms of human T-lymphotropic virus type 1 transmission and disease. Curr. Opin. Virol. 2, 474–481. http://dx.doi.org/10.1016/j.coviro.2012.06.007.
Maloney, E.M., Cleghorn, F.R., Morgan, O.S., Rodgers-Johnson, P., Cranston, B., Jack, N., Blattner, W.A., Bartholomew, C., Manns, A., 1998. Incidence of HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) in Jamaica and Trinidad. J. Acquir. Immune Defic. Syndr. Hum. Retrovirology Off. Publ. Int. Retrovirology Assoc. 17, 167–170.
Matsuoka, M., 2010. HTLV-1 bZIP factor gene: Its roles in HTLV-1 pathogenesis. Mol. Aspects Med. Role of human retroviruses (HIVs and HTLVs) non structural proteins in virus host interaction 31, pp. 359–366. http://dx.doi.org/10.1016/j.mam.2010.06.002.
Matsuoka, M., Green, P.L., 2009. The HBZ gene, a key player in HTLV-1 pathogenesis. Retrovirology 6, 71. http://dx.doi.org/10.1186/1742-4690-6-71.
Matsuzaki, T., Nakagawa, M., Nagai, M., Usuku, K., Higuchi, I., Arimura, K., Kubota, H., Izumo, S., Akiba, S., Osame, M., 2001. HTLV-I proviral load correlates with progression of motor disability in HAM/TSP: analysis of 239 HAM/TSP patients including 64 patients followed up for 10 years. J. Neurobiol. 7, 228–234. http://dx.doi.org/10.1080/13550280152403272.
Nagai, M., Usuku, K., Matsumoto, W., Kodama, D., Takenouchi, N., Moritoyo, T., Hashiguchi, S., Ichinose, M., Bangham, C.R., Izumo, S., Osame, M., 1998. Analysis of HTLV-I proviral load in 202 HAM/TSP patients and 243 asymptomatic HTLV-I carriers: high proviral load strongly predisposes to HAM/TSP. J. Neurobiol. 4, 586–593.
Pineda-Pena, A.-C., Schrooten, Y., Vinken, L., Ferreira, F., Li, G., Trovao, N.S., Khouri, R., Derdelinckx, I., De Munter, P., Kucherer, C., Kostrikis, L.G., Nielsen, C., Littsola, K., Wensing, A., Stanojevic, M., Paredes, R., Balotta, C., Albert, J., Boucher, C., Gomez-Lopez, A., Van Wijngaerden, E., Van Ranst, M., Vercauteren, J., Vandamme, A.-M., Van Laethem, K., 2014. Trends and predictors of transmitted drug resistance (TDR) and clusters with TDR in a local Belgian HIV-1 epidemic. PLoS One 9, e101738. http://dx.doi.org/10.1371/journal.pone.0101738.
Pise-Masison, C.A., de Castro-Amarante, M.F., Enose-Akahata, Y., Buchmann, R.C., Fenizia, C., Washington Parks, R., Edwards, D., Fiocchi, M., Alcantara, L.C., Bialuk, I., Graham, J., Walser, J.-C., McKinnon, K., Galvão-Castro, B., Gessain, A., Venzon, D., Jacobson, S., Franchini, G., 2014. Co-dependence of HTLV-1 p12 and p8 functions in virus persistence. PLoS Pathog. 10. http://dx.doi.org/10.1371/journal.ppat.1004454.
Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., Gallo, R.C., 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. 77, 7415–7419.
Prosperi, M.C.F., De Luca, A., 2012. Computational models for prediction of response to antiretroviral therapies. AIDS Rev. 14, 145–153.
Trovato, R., Mulloy, J.C., Johnson, J.M., Takemoto, S., de Oliveira, M.P., Franchini, G., 1999. A lysine-to-arginine change found in natural alleles of the human T-cell lymphotropic/leukemia virus type 1 p12(I) protein greatly influences its stability. J. Virol. 73, 6460–6467.
Van Prooyen, N., Gold, H., Andresen, V., Schwartz, O., Jones, K., Ruscetti, F., Lockett, S., Gudla, P., Venzon, D., Franchini, G., 2010. Human T-cell leukemia virus type 1 p8 protein increases cellular conduits and virus transmission. Proc. Natl. Acad. Sci. U. S. A. 107, 20738–20743. http://dx.doi.org/10.1073/pnas.1009635107.
WiaF, E., 2011. Data Mining: Practical Machine Learning Tools and Techniques.
Yamano, Y., Nagai, M., Brennan, M., Mora, C.A., Soldan, S.S., Tomaru, U., Takenouchi, N., Izumo, S., Osame, M., Jacobson, S., 2002. Correlation of human T-cell lymphotropic virus type 1 (HTLV-1) mRNA with proviral DNA load, virus-specific CD8(+) T cells, and disease severity in. Blood 99, 88–94.
Yoshida, M., Miyoshi, I., Hinuma, Y., 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc. Natl. Acad. Sci. U. S. A. 79, 2031–2035.

4.3 Assessment of genetic diversity of HTLV-1 ORF-I sequences collected from patients with different clinical profiles.

Assessment of genetic diversity of HTLV-1 ORF-I sequences collected from patients with different clinical profiles

Fernanda Khouri Barreto[a]; Bernardo Galvão-Castro[b]; Lourdes Farre[a]; Achilea Candida Lisboa Bittencourt[c]; Melina Mosquera Borba[a]; Thessika Hialla Almeida Araújo[a,b]; Filipe Ferreira de Almeida Rego[d]; Luiz Carlos Junior Alcantara[a,*]; Maria Fernanda de Castro-Amarante[a,*]

[a] Fundação Oswaldo Cruz - Centro de Pesquisa Gonçalo Moniz, Salvador, Bahia, Brazil

[b] Escola Bahiana de Medicina e Saúde Pública, Salvador, Bahia , Brazil

[c] Universidade Federal da Bahia, Brazil

[d] Universidade Católica do Salvador, Salvador, Bahia, Brazil

[*]Corresponding author

E-mail: lalcan@bahia.fiocruz.br

E-mail: mfamarante@gmail.com

Abstract

The human T-cell lymphotropic virus type 1 (HTLV-1) infects 5 to 10 million individuals and remains without specific treatment. This retrovirus is the etiologic agent of infective dermatitis associated to HTLV-1 (IDH), HTLV-1-associated myelopathy/ tropical spastic paraparesis (HAM/TSP), adult T-cell leukaemia/lymphoma (ATLL), among others. The HTLV-1 genome is composed of the genes gag, pol, env and a region near the 3' end, known as pX. This region contains four open reading frames (ORFs) that encode specific regulatory proteins. The ORF-I produces the protein p12 and its cleavage product p8. Here, we analyzed the genetic diversity of 63 ORF-I sequences from patients with ATLL, HAM/TSP, IDH and asymptomatic individuals. Six natural amino acid changes with frequency over 5% were identified: G29S, P34L, F61L, S63P, F78L and S91P. Among these mutations, only P34L was found with a statistically significant difference in the frequency within the IDH and HAM/TSP groups (p=0.047). Also, P34L mutation was the only one able to alter protein profile. The identification of regions where the post-translational sites were identified also showed a high identity among the sequences and the amino acid changes exclusive of specific clinical profile were found in less than 5% of the samples. The low overall genetic diversity found in ORF-I sequences from patients with different clinical profiles (0.007) corroborates the fact that HTLV-1 genome exhibits relatively few sequence variations and suggested that this region could be used in the HTLV-1 therapeutic vaccine development.

Keywords: HTLV-1; ORF-I; HAM/TSP; ATLL; IDH

1. Introduction

The human T-cell lymphotropic virus type 1 (HTLV-1) was the first described human retrovirus [1]. It is estimated that 5-10 million people are infected with HTLV-1 in the world and although this infection is endemic in different geographic regions it still remains without effective therapeutic methods [2].

The HTLV-1 genome is composed of genes gag, pol, env, and a region near to the 3' end, known as pX. The pX region contains four open reading frames (ORFs) that encode specific regulatory proteins [3,4]. The ORF-I encodes the p12 protein that resides in the endoplasmic reticulum (ER). This protein is susceptible to two protein cleavages. The first cleavage occurs between amino acids 9 and 10 and results in the loss of the non-canonical endoplasmic reticulum retention signal. The second cleavage occurs between amino acids 29 and 30 and is responsible for p8 protein release, that moves to the infected cell surface [5]. The ORF-II encodes the proteins p30 and p13, while ORF-III and ORF-IV produce the proteins Rex and Tax, respectively [6,7].

It is known that most patients infected with HTLV-1 do not develop clinical manifestations, however this retrovirus is the etiologic agent of inflammatory diseases as infective dermatitis associated to HTLV-1 (IDH) and HTLV-1-associated myelopathy/ tropical spastic paraparesis (HAM/TSP), among others. HTLV-1 is also responsible for the development of an aggressive clonal disease known as adult T-cell leukaemia/lymphoma (ATLL), [8–10].

The major challenge in HTLV-1 infection is to understand why some individuals develop pathological processes while others remain asymptomatic and what is the best way to prevent viral persistence and infectivity. Additionally, HTLV-1-infected subjects remain under clinical latency for years to decades, which interfere with diagnostic and treatment during early infection.

A recent research demonstrated that the persistence of HTLV-1 infection is influenced to the expression of p12 and p8 proteins, encoded from the ORF-I of the pX gene region. This study suggest that some natural ORF-I mutations alter the expression of the p12 and p8 proteins and that equivalent amounts of both are necessary to prevent recognition and lysis of HTLV-1 cells infected by Cytotoxic T Cells (CTLs) [11]. Furthermore, Barreto et al., using bioinformatics analysis suggested that some of these natural ORF-I mutations might influence the proviral load and clinical manifestation of HAM/TSP [12].

The p12 and p8 proteins have been found to interfere in several intracellular mechanisms. Banerjee et al. demonstrated that by inhibiting histocompatibility complex class I (MHC-I) expression, p12 decreases the presentation of viral peptides to CTLs. At the same time, p12 also decrease the expression of ICAM-1 and ICAM-2 in T cells, thereby inhibiting adhesion and recognition of HTLV-1-infected T cells by NK cells [13].

The literature also shows that p12 protein increase the intracellular calcium concentration by inducing calcium release from the ER through inositol trisphosphate receptors and increasing the Ca^{2+} influx through the plasma membrane [14]. The increased intracellular Ca^{2+} concentration impacts on several cellular mechanisms, which in turn work on behalf of HTLV-1 infection, as the regulation of viral transcription through the modulation of the cofactor p300 or viral propagation through the activation of the nuclear factor of activated T-cells (NFAT) [15–17]. In the lipid rafts at the plasma membrane, p8 protein mediates viral spread by inducing the clustering of lymphocyte function-associated antigen 1(LFA-1) molecules and the formation of cellular conduits that cooperate to viral transmission [17,18]. In addition, p8 is able to inhibit TCR signalling causing T cell anergy [19].

Considering the critical role of p12 and p8 proteins on the course of infection, this study aims to evaluate whether HTLV-1 ORF-1 region could be used as a target for the

development of a therapeutic vaccine analyzing the genetic diversity of ORF-1 among patients with ATLL, HAM/TSP, IDH, and asymptomatic individuals.

2. Materials and methods

To evaluate the genetic characteristics of HTLV-1 ORF-I sequences, blood samples of 63 patients with defined clinical profiles were collected: 16 samples from patients with HAM/TSP, 13 samples from ATLL patients, 23 samples from asymptomatic patients and 11 samples from patients with IDH. The clinical classification was carried out by medical experts according to World Health Organization (WHO). All samples were anonymized and informed consent was written and obtained from each subject at the Center of HTLV-1 Bahia School of Medicine and Public Health and the Hospital Complex Prof. Edgar Santos, both located in Salvador City, Bahia State, and at the Fundação Hemocentro de Ribeirão Preto (Fundherp) located in Ribeirão Preto City, São Paulo State. The study was approved by the Ethics Committee of the Gonçalo Moniz Institute (IGM), Fiocruz.

2.1 ORF-I PCR and sequencing

The peripheral blood mononuclear cells (PBMCs) were obtained from EDTA blood by density gradient centrifugation. DNA was extracted using spin column DNA extraction system (QIAamp DNA Mini Kit - Qiagen) and the samples were used for amplification of ORF-I through Polymerase Chain Reaction (PCR). The PCR conditions were as follows: denaturation at 94°C for 3 minutes, annealing at 94°C (15 sec), 65°C (45 sec), 72°C (1 minute), 35 times of cycle, and a final extension of 72°C for 8 minutes, with the primers 24^{+} (5'CGTATCGCCTCCCTCGCGCCATCAGAGTATGCTGCCCAGAACAG3') and 27^{-} (5'CTATGCGCCTTGCCAGCCCGCTCAGGGTTCCATGTATCCATTTCGGA3'). The 400 base pairs products were analysed on 1% agarose gel. The PCR products were submitted to a different sequencing methods: Sanger and Ion Torrent. For Sanger

sequencing, 32 amplicons were purified using PureLink PCR Purification Kit (Thermo Fisher Scientific) and sequenced in an ABI Prism 3100 DNA Sequencer (Applied Biosystems Inc., Foster City, CA) using Taq FS Dye (Applied Biosystems) terminator cycle sequencing with the same PCR primers. The other 31 samples were sequenced using the Ion Torrent technology (Life Technologies, Carlsbad, CA). For Ion Torrent sequencing, the amplicons of each patient were fragmented using the Bioruptor standard (Diagenode). The libraries were built by the automated AB Library Builder System (Applied Biosystems). To normalize the number of molecules required for the emulsion PCR, a quantitation step was performed using the automated Ion OneTouch 2 platform. Ion Xpress Adapters and unique Ion Xpress Barcodes (Life Technologies) were linked to fragmented material by using the Ion Plus Fragment Library kit, followed by the Emulsion PCR (ePCR) and enrichment steps that were performed using Ion PGM OT2 200 Template Kit (Thermo Fisher Scientific) according to the manufacturer's protocol (Ion 2 platform OT2 OneTouch System - Life Technologies). Finally, the mix was loaded on a 318v2 chip and the sequencing reaction was performed on the PGM device.

2.2 Dataset construction and analysis

The fastaq files from the 31 sequences generated by Ion Torrent technology was trimmed, manual edited and aligned to the HTLV-1 reference sequence ATK-1 (J02029) to generate the consensus sequence of each patient. The 31 consensus sequences were aligned to the 32 sequences generated by Sanger sequencing in order to determine the mutations profile. The final dataset was composed for 63 HTLV-1 ORF-I complete sequences and submitted to a search for the major and minor natural amino acid changes. All these analysis were done using Geneious R6 software [20]. The statistical analyses were performed using Fisher's exact test and P values lower than 0.05 were considered statistically significant.

To identify the presence or absence of the major mutations in viral quasispecies, the reads of Ion Torrent sequencing were analyzed. First, the adaptor contamination was removed from the reads and a consensus reference sequence was generated. The sequencing reads were trimmed from both sides and mapped to this consensus reference sequence using the lowest sensitivity (maximum mismatches per read: 10%). These analyses were done using the Geneious Software [20].

2.3 Assessment of molecular characteristics

To perform the molecular analysis of the identified mutations, physico-chemical analysis was carried out using Network Protein Sequence Analysis (NPS@) (http://npsa-pbil.ibcp.fr/) and the potential protein domain analysis was executed using the Prosite tool [21,22]. The genetic distances within and between the sequences from patients with different clinical profiles were measured using the Mega 6.0 program [23].

3. Results

We analyzed 63 ORF-I sequences from HTLV-1 patients with the following clinical profiles: 16 samples from patients with HAM/TSP, 13 samples from ATLL patients, 23 samples from asymptomatic patients and 11 samples from patients with IDH. The mutations found in the dataset were described as minor (if the mutation prevalence was lower than 5%) or major (if the mutation prevalence was higher or equal to 5%).

3.1 Evaluation of the major frequent ORF-I natural mutations

Six natural amino acid changes with frequency over 5% were identified within the dataset: G29S, P34L, F61L, S63P, F78L and S91P. Among them, five were located in specific motifs and were previously described as mutations that influence the expression profile of HTLV-1 ORF-I protein products: G29S, P34L, F61L, S63P and S91P [11]. Among the six mutations identified, only P34L was found with a statistically significant difference in the frequency within the IDH and HAM/TSP groups ($p = 0.047$) (Table 1).

175 The search of these six major amino acid changes in viral quasispecies revealed that S63P
176 and S91P mutations appear in high frequency among the reads of each patient, while the
177 other mutations were found both in high and low frequency.
178 The wild type and mutated sequences were submitted to physico-chemical analysis and
179 only the P34L mutation was able to alter protein profile. The NPS@ analysis suggested
180 that the ORF-I product with a leucine in 34 position was less hydrophilic, flexible and
181 antigenic than the wild type. The accessibility was also decreased while the hydropathy
182 and membrane buried-helix profile were slightly increased (Figure 1).
183 To identify if these amino acid mutations were able to create or abrogate potential protein
184 domains we submitted the 63 ORF-I sequences to the scan prosite tool and no changes
185 were observed. All sequences have a casein kinase II phosphorylation site at the 23-26
186 position and a protein kinase C phosphorylation site at the 75-77 position, which were not
187 altered by the mutations (data not shown).
188 ***3.2 ORF-I genetic diversity from patients with HAM/TSP, ATLL, HID and***
189 ***asymptomatic***
190 The analysis of ORF-I sequences revealed 13 mutations found in less than 5% of the
191 samples. Despite being in low frequency all these mutations have an important
192 characteristic: they are observed only in specific clinical profiles. Six amino acid changes
193 were detected in samples from asymptomatic patients (L5I, S7G, P73S, R82*, A96V and
194 F99L), five mutations were exclusive of HAM/TSP sequences (G25S, C39Y, F54L, P86S
195 and A89T), while L55F and F84L mutation was identified only in IDH and ATLL
196 samples, respectively.
197 The overall diversity between sequences from patients with HAM/TSP, ATLL, IDH and
198 asymptomatic was 0.007 and the genetic distance values within and between the different
199 clinical profiles were described in Table 2.

4. Discussion

The HTLV-1 was discovered more than 35 years ago, infects 5 to 10 million individuals worldwide and remains without specific treatment [1,2]. In most cases the severity of the HTLV-associated diseases is influenced by the high proviral load, through clonal expansion of infected cells rather than to the use of the viral reverse transcriptase [24]. Given the genetic stability of HTLV-1 it was originally believed that developing an anti-HTLV-1 prophylactic vaccine would be an easy task [25]. However studies demonstrated that the induction of HTLV-1 protective immune response is not so simple [25–28]. Here we suggest that a therapeutic vaccine may be a better alternative and the HTLV-1 ORF-I is a good target for the development of this vaccine.

Previous studies shows that the expression of p12 and p8 proteins can influence the efficient viral persistence and the outcome of HTLV-1 infection [11,12]. These data suggest that the inhibition of p12 cleavages enhances the cytotoxic cells ability to recognize cells infected, thereby decreasing the proviral load and the HTLV-1 persistence. Therefore, whether p12 cleavages can be inhibited perhaps HTLV-1 infection may be controlled.

The analysis of ORF-I sequences from patients with ATLL, HAM/TSP, IDH and asymptomatic individuals demonstrated the low diversity of this region and reinforce that this region may be a good target for HTLV-1 treatment. The frequency evaluation of ORF-I mutations from patients with different clinical profiles reveals that only one mutation was found with a statistically significant between IDH and HAM/TSP groups (P34L). The identification of regions where the post-translational sites were identified also showed a high identity among the sequences and the amino acid changes exclusive of specific clinical profile were found in less than 5% of the samples.

224 The low overall genetic diversity found in ORF-I sequences from patients with different
225 clinical profiles corroborates the fact that the HTLV-1 genome exhibits relatively few
226 sequence variations and that the development of a therapeutic vaccine is possible.

227 ## 5. Conclusions

228 The present study demonstrated the low diversity of HTLV-1 ORF-I sequences from
229 patients with ATLL, HAM/TSP, IDH and asymptomatic individuals. These data
230 suggested that this region is a possible target for a therapeutic vaccine that could be used
231 in patients with different clinical profiles. More analyses involving sequences from
232 patient with other HTLV-1 pathologies can provide more information about the ORF-I
233 genetic diversity and these data can be used to design an HTLV-1 vaccine.

234 ## Availability of supporting data

235 All sequences are available in the GenBank database (accession numbers KY007244-
236 KY007274; MF158987-MF159019).

237 ## Competing interests

238 The authors declare that they have no competing interests.

239 ## Acknowledgements

240 Not applicable.

241 ## References

242 1. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and
243 isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient
244 with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. 1980;77:7415–9.
245 2. Gessain A, Cassar O. Epidemiological Aspects and World Distribution of HTLV-1
246 Infection. Front. Microbiol. 2012;3:388.
247 3. Berneman ZN, Gartenhaus RB, Reitz MS, Blattner WA, Manns A, Hanchard B, et al.

248 Expression of alternatively spliced human T-lymphotropic virus type I pX mRNA in
249 infected cell lines and in primary uncultured cells from patients with adult T-cell
250 leukemia/lymphoma and healthy carriers. Proc. Natl. Acad. Sci. U. S. A. 1992;89:3005–
251 9.
252 4. Ciminale V, Pavlakis GN, Derse D, Cunningham CP, Felber BK. Complex splicing in
253 the human T-cell leukemia virus (HTLV) family of retroviruses: novel mRNAs and
254 proteins produced by HTLV type I. J. Virol. 1992;66:1737–45.
255 5. Koralnik IJ, Gessain A, Klotman ME, Lo Monico A, Berneman ZN, Franchini G.
256 Protein isoforms encoded by the pX region of human T-cell leukemia/lymphotropic virus
257 type I. Proc. Natl. Acad. Sci. U. S. A. 1992;89:8813–7.
258 6. Kiyokawa T, Seiki M, Iwashita S, Imagawa K, Shimizu F, Yoshida M. p27x-III and
259 p21x-III, proteins encoded by the pX sequence of human T-cell leukemia virus type I.
260 Proc. Natl. Acad. Sci. U. S. A. 1985;82:8359–63.
261 7. Kiyokawa T, Seiki M, Imagawa K, Shimizu F, Yoshida M. Identification of a protein
262 (p40x) encoded by a unique sequence pX of human. Gan. 1984;75:747–51.
263 8. La Grenade L, Manns A, Fletcher V, Derm D, Carberry C, Hanchard B, et al. Clinical,
264 pathologic, and immunologic features of human T-lymphotrophic virus type I-associated
265 infective dermatitis in children. Arch. Dermatol. 1998;134:439–44.
266 9. Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, et al. Antibodies to
267 human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet
268 Lond. Engl. 1985;2:407–10.
269 10. Yoshida M, Miyoshi I, Hinuma Y. Isolation and characterization of retrovirus from
270 cell lines of human adult T-cell leukemia and its implication in the disease. Proc. Natl.
271 Acad. Sci. U. S. A. 1982;79:2031–5.
272 11. Pise-Masison CA, de Castro-Amarante MF, Enose-Akahata Y, Buchmann RC,

273 Fenizia C, Washington Parks R, et al. Co-dependence of HTLV-1 p12 and p8 Functions
274 in Virus Persistence. PLoS Pathog. [Internet]. 2014 [cited 2015 Jul 10];10. Available
275 from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4223054/
276 12. Barreto FK, Khouri R, Rego FF de A, Santos LA, Castro-Amarante MF de, Bialuk I,
277 et al. Analyses of HTLV-1 sequences suggest interaction between ORF-I mutations and
278 HAM/TSP outcome. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.
279 2016;45:420–5.
280 13. Banerjee P, Feuer G, Barker E. Human T-Cell Leukemia Virus Type 1 (HTLV-1)
281 p12I Down-Modulates ICAM-1 and -2 and Reduces Adherence of Natural Killer Cells,
282 Thereby Protecting HTLV-1-Infected Primary CD4+ T Cells from Autologous Natural
283 Killer Cell-Mediated Cytotoxicity despite the Reduction of Major Histocompatibility
284 Complex Class I Molecules on Infected Cells. J. Virol. 2007;81:9707–17.
285 14. Fukumoto R, Andresen V, Bialuk I, Cecchinato V, Walser J-C, Valeri VW, et al. In
286 vivo genetic mutations define predominant functions of the human T-cell
287 leukemia/lymphoma virus p12I protein. Blood. 2009;113:3726–34.
288 15. Nair AM, Michael B, Datta A, Fernandez S, Lairmore MD. Calcium-dependent
289 enhancement of transcription of p300 by human T-lymphotropic type 1 p12I. Virology.
290 2006;353:247–57.
291 16. Ding W, Albrecht B, Kelley RE, Muthusamy N, Kim S-J, Altschuld RA, et al. Human
292 T-cell lymphotropic virus type 1 p12(I) expression increases cytoplasmic calcium to
293 enhance the activation of nuclear factor of activated T cells. J. Virol. 2002;76:10374–82.
294 17. Kim S-J, Nair AM, Fernandez S, Mathes L, Lairmore MD. Enhancement of LFA-1-
295 mediated T cell adhesion by human T lymphotropic virus type 1 p12I1. J. Immunol.
296 Baltim. Md 1950. 2006;176:5463–70.
297 18. Van Prooyen N, Gold H, Andresen V, Schwartz O, Jones K, Ruscetti F, et al. Human

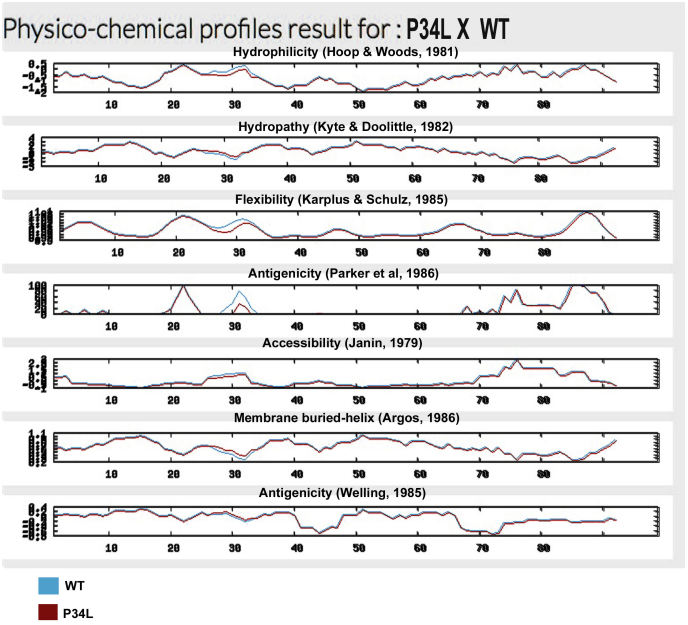
298 T-cell leukemia virus type 1 p8 protein increases cellular conduits and virus transmission.
299 Proc. Natl. Acad. Sci. U. S. A. 2010;107:20738–43.
300 19. Rosadas C, Puccioni-Sohler M. HTLV-1 ORF-I Encoded Proteins and the Regulation
301 of Host Immune Response: Viral Induced Dysregulation of Intracellular Signaling. J.
302 Immunol. Res. 2015;2015:498054.
303 20. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious
304 Basic: an integrated and extendable desktop software platform for the organization and
305 analysis of sequence data. Bioinforma. Oxf. Engl. 2012;28:1647–9.
306 21. Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: network protein sequence
307 analysis. Trends Biochem. Sci. 2000;25:147–50.
308 22. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and
309 continuing developments at PROSITE. Nucleic Acids Res. 2013;41:D344-347.
310 23. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular
311 Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 2013;30:2725–9.
312 24. Wattel E, Vartanian JP, Pannetier C, Wain-Hobson S. Clonal expansion of human T-
313 cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers
314 without malignancy. J. Virol. 1995;69:2863–8.
315 25. de Thé G, Bomford R. An HTLV-I vaccine: why, how, for whom? AIDS Res. Hum.
316 Retroviruses. 1993;9:381–6.
317 26. Bomford R, Kazanji M, De Thé G. Vaccine against human T cell leukemia-
318 lymphoma virus type I: progress and prospects. AIDS Res. Hum. Retroviruses.
319 1996;12:403–5.
320 27. Rodríguez SM, Florins A, Gillet N, de Brogniez A, Sánchez-Alcaraz MT, Boxus M,
321 et al. Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV.
322 Viruses. 2011;3:1210–48.

323 28. Mahieux R. A vaccine against HTLV-1 HBZ makes sense. Blood. 2015;126:1052–3.

324

Figures

Figure 1 - Physico-chemical analysis of P34L mutation versus wild type.



The graphs are organized as follows: Hydrophilicity; Hydropathy; Flexibility; Antigenicity; Accessibility; Membrane buried-helix; Antigenicity.

Tables

Table 1 – Frequency of major ORF-I natural mutations and their respective motif.

Mutation	IDH (n=11)	HAM/TSP (n=15)	ATLL (n=13)	Asymptomatic (n=23)	Motif
G29S	-	2	1	2	Transmembrane domain
P34L[†, *]	2	10	4	8	SH3 binding domain
F61L	-	2	1	1	Transmembrane domain
S63P	9	14	12	19	Transmembrane domain
F78L	-	-	1	2	-
S91P	10	14	12	19	SH3 binding domain

([†]) Mutation able to change the chemical physical profile; (*) p = 0.047 between IDH and HAM/TSP profiles. SH3 = Src homology 3.

Table 2 - Genetic distances in HTLV-1 ORF-I sequences from patients with different clinical profiles.

	IDH	HAM/TSP	ATLL	Asymptomatic
IDH	0.006	0.007	0.006	0.007
HAM/TSP	0.007	0.007	0.007	0.008
ATLL	0.006	0.007	0.006	0.007
Asymptomatic	0.007	0.008	0.007	0.008

5 DISCUSSÃO

O HTLV-1 infecta aproximadamente 5-10 milhões de pessoas no mundo e é o agente etiológico de diversas patologias, como a leucemia/linfoma de células T do adulto (ATLL), a paraparesia espástica tropical/mielopatia associada ao HTLV (HAM/TSP), a dermatite infectiva associada ao HTLV-1 (DIH), entre outras (YOSHIDA et al., 1982; GESSAIN, A. et al., 1985; GONÇALVES et al., 2003; GESSAIN, ANTOINE e CASSAR, 2012).

Os indivíduos infectados pelo HTLV-1 podem permanecer assintomáticos ou desenvolver manifestações clínicas e a gravidade das doenças pode estar relacionada, dentre outros fatores, à alta carga proviral (MATSUZAKI et al., 2001; YAMANO et al., 2006). Dado ao fato do genoma desse retrovírus ser considerado estável, linhas de pesquisa acreditavam que o desenvolvimento de uma vacina profilática contra a infecção seria fácil, mas estudos demonstraram que a indução de uma resposta imune protetora contra a infecção pelo HTLV-1 não é tão simples (DE THÉ e BOMFORD, 1993; BOMFORD, KAZANJI e DE THÉ, 1996; RODRÍGUEZ et al., 2011; MAHIEUX, 2015). Nesta tese, avaliamos se a ORF-I da região pX do HTLV-1 poderia ser utilizada como alvo para o desenvolvimento de uma vacina terapêutica. Os dados obtidos demonstram a baixa diversidade genética dessa região, corroborando com a hipótese desta tese de que a alteração da clivagem da p12 poderia ser utilizada para o controle da carga proviral.

Em 2014, Pise-Masison e colaboradores identificaram mutações naturais na ORF-I capazes de alterar a o padrão de clivagem dessa região, aumentando a concentração de p12 ou de p8. A partir desse estudo foi demonstrada que a co-expressão de p12 e p8 é importante para a persistência e infectividade viral (PISE-MASISON et al., 2014). Esses dados sugerem que a inibição da clivagem da p12 aumenta a capacidade de células T citotóxicas reconheceram células infectadas e diminuem a capacidade de infecção de células-alvo. Dessa forma, essa estratégia poderia ser utilizada para induzir a diminuição da carga proviral e o controle da infecção pelo HTLV-1.

Para avaliar se a estratégia de desenvolver uma vacina terapêutica baseada na inibição da clivagem da p12 poderia ser utilizada em diversas manifestações clínicas associadas à infecção pelo HTLV-1, sequências da ORF-I provenientes de pacientes com HAM/TSP, ATLL, DIH e indivíduos assintomáticos foram analisadas na presente tese.

Caracterizações anteriores da ORF-I do HTLV-1 identificaram a existência de diversos domínios funcionais, incluindo um sinal de retenção no retículo endoplasmático, dois domínios transmembrana e dois *leucine zipper* (LZ), um sítio de ligação para a calcineurina e quatro regiões SH3 (*proline-rich Src homology*) (DING et al., 2002; FUKUMOTO et al., 2009). Algumas mutações identificadas e analisadas nesta tese estão localizadas em domínios funcionais.

As mutações D26N e F61L estão localizadas no primeiro e segundo domínio transmembrana, respectivamente, e foram observadas em maior frequência em pacientes com baixa carga proviral. Os domínios transmembrana são importantes para manter as proteínas na sua conformação e localização corretas e quando alterados podem influenciar a função proteica e, consequentemente, o curso da infecção.

A mutação R83C, também identificada em maior frequência em pacientes com baixa carga proviral, encontra-se no sítio de ligação para a calcineurina. A calcineurina participa de uma via importante na ativação de células T, que implica na desfosforilação do fator nuclear de ativação de células T (NFAT). Esse fator é essencial para a ativação da expressão de interleucina-2 (IL-2), a qual exerce papel importante na proliferação de células T. Contrariamente, a p12 é capaz de diminuir a ativação dessas células ao se ligar à calcineurina e competir com o NFAT, impactando na resposta imune e replicação viral (ALBRECHT e LAIRMORE, 2002; KIM et al., 2003). Essa mutação pode, portanto, alterar a interação entre a p12 e a calcineurina, influenciando no desfecho da infecção.

Sítios de modificação pós-traducionais também podem impactar na conformação das proteínas e, consequentemente, na função das mesmas. A identificação de sítios de modificação pós-traducionais mostrou a elevada identidade entre as sequências provenientes de pacientes com ATLL, HAM/TSP, DIH e assintomáticos. Todas as sequências apresentaram os mesmo dois únicos sítios: sítio de fosforilação para a kinase II na posição 23-26 e sítio de fosforilação para a kinase C, na posição 75-77.

Apesar de mutações naturais na ORF-I serem observadas em todos os perfis clínicos avaliados, mutações não-sinônimas exclusivas de determinado perfil clínico foram encontradas em menos de 5% das sequências analisadas, demonstrando o alto grau de conservação dessa região gênica. A diversidade genética global da ORF-I também corrobora com o fato do genoma do HTLV-1 ser considerado estável e reforça a hipótese de que essa região pode ser um bom alvo para o desenvolvimento de uma terapia contra essa infecção.

Todas essas análises somente foram possíveis após identificar a posição nucleotídica da ORF-I no genoma viral. Durante essa etapa foi possível verificar que o principal genoma completo do HTLV-1 (ATK1) não estava completamente anotado no GenBank. A determinação exata da localização dos genes de um protótipo amplamente utilizado em análises que envolvem estudos de variabilidade e de produção de proteínas virais colabora de forma positiva para investigações futuras que busquem responder questões que ainda permanecem sem resposta na infecção pelo HTLV-1. Dentro desse contexto, a anotação completa do genoma do HTLV-1 realizada nesta tese pode servir de base para novos estudos.

6 CONCLUSÃO

Com a caracterização molecular da ORF-I proveniente de pacientes com diferentes perfis clínicos foi possível evidenciar a baixa diversidade genética da ORF-I do HTLV-1, além de demonstrar que o desenvolvimento de HAM/TSP pode estar relacionado à condição de dependência entre mutações nessa região gênica e a alta carga proviral. Esses dados sugerem que a ORF-I do HTLV-1 poderia ser utilizada como alvo no desenvolvimento de uma vacina terapêutica.

REFERÊNCIAS

ALBRECHT, B. et al. Activation of nuclear factor of activated t cells by human t-lymphotropic virus type 1 accessory protein p12(I). **Journal of Virology,** v. 76, n. 7, p. 3493-3501, 2002.

AMORIM, C. F. et al. Functional activity of monocytes and macrophages in htlv-1 infected subjects. **PLoS Neglected Tropical Diseases,** v. 8, n. 12, p. e3399, 2014.

BANERJEE, P.; FEUER, G.; BARKER, E. Human t-cell leukemia virus type 1 (htlv-1) p12i down-modulates icam-1 and -2 and reduces adherence of natural killer cells, thereby protecting htlv-1-infected primary cd4+ t cells from autologous natural killer cell-mediated cytotoxicity despite the reduction of major histocompatibility complex class i molecules on infected cells. **Journal of Virology,** v. 81, n. 18, p. 9707-9717, 2007.

BANGHAM, C. R. M. The immune control and cell-to-cell spread of human t-lymphotropic virus type 1. **Journal of General Virology,** v. 84, n. 12, p. 3177-3189, 2003.

BARBEAU, B.; MESNARD, J-M. Making sense out of antisense transcription in human t-cell lymphotropic viruses (Htlvs). **Viruses,** v. 3, n. 12, p. 456-468, 2011.

BINDHU, M.; NAIR, A. LAIRMORE, M. D. Role of accessory proteins of htlv-1 in viral replication, t cell activation, and cellular gene expression. **Front Bioscience,** v. 9, n. p. 2556-2576, 2004.

BITTENCOURT, A. L. et al. Adult t-cell leukemia/lymphoma in bahia, brazil: analysis of prognostic factors in a group of 70 patients. **American Journal of Clinical Pathology**. v. 128, n. 5, p. 875-882, 2007.

BOMFORD, R.; KAZANJI, M.; DE THÉ, G. Vaccine against human T cell leukemia-lymphoma virus type I: progress and prospects. **AIDS Research Human Retroviruses,** v. 12, n. 5, p. 403-405, 1996.

BRADY, J. et al. Identification of P4ox-responsive regulatory sequences within the human T-Cell leukemia virus type i long terminal repeat. **Journal of Virology**. v. 61, n. 7, p. 2175-2181, 1987.

CARNEIRO-PROIETTI, A. B. et al. Infection and disease caused by the human t cell lymphotropic viruses type i and ii in brazil. **Revista da Sociedade Brasileira de Medicina Tropical,** v. 35, n. 5, p. 499-508, 2002.

CATALAN-SOARES, B.; CARNEIRO-PROIETII, A. B.; PROIETTI, F. A. Grupo interdisciplinar de pesquisas em htlv. vírus-t linfotrópico humano em familiares de candidatos a doação de sangue soropositivos: disseminação silenciosa. **Revista Panamericana Salud Publica**. v. 16, n. 6, p. 387-394, 2004.

CATALAN-SOARES, B.; CARNEIRO-PROIETTI, A. B.; PROIETTI, F. A. Heterogeneous geographic distribution of human t-cell lymphotropic viruses i and ii

(htlv-i/ii): serological screening prevalence rates in blood donors from large urban areas in Brazil. **Cadernos de Saude Publica,** v. 21, n. 3, p. 926-931, 2005.

CIMARELLI, A. et al. Clonal expansion of human t-cell leukemia virus type ii in patients with high proviral load. **Virology,** v. 223, n. 2, p. 362-364, 1996.

COOK, L. B. et al. Htlv-1: Persistence and Pathogenesis. **Virology,** v. 435, n. 1, p. 131-140, 2013.

COURGNAUD, V. et al. Simian t-cell leukemia virus (stlv) infection in wild primate populations in cameroon: evidence for dual stlv type 1 and type 3 infection in Agile Mangabeys (Cercocebus Agilis). **Journal of Virology,** v. 78, n. 9, p. 4700-4709, 2004.

DELAMARRE, L. et al. The HTLV-I envelope glycoproteins: structure and functions. **Journal of Acquired Immune Deficience Syndrome Human Retrovirology,** v. 13 Suppl 1, n. p. S85-91, 1996.

DE THÉ, G.; BOMFORD, R. An HTLV-I vaccine: why, how, for whom? **AIDS Research Human Retroviruses,** v. 9, n. 5, p. 381-866, 1993.

DING, W. et al. Human T-Cell Lymphotropic Virus Type 1 p12(i) expression increases cytoplasmic calcium to enhance the activation of nuclear factor of activated t cells. **Journal of Virology,** v. 76, n. 20, p. 10374-10382, 2002.

DOURADO, I. et al. HTLV-I in the general population of salvador, brazil: a city with african ethnic and sociodemographic characteristics. **Journal of Acquired Immune Deficience Syndrome,** v. 34, n. 5, p. 527-531, 2003.

EDWARDS, D. et al. ORF-I AND ORF-II-encoded proteins in htlv-1 infection and persistence. **Viruses,** v. 3, n. 6, p. 861-885, 2011.

EGUCHI, K. et al. Primary Sjogren's Syndrome with antibodies to HTLV-I: clinical and laboratory features. **Annals of the Rheumatic Diseases,** v. 51, n. 6, p. 769-776, 1992.

FRANCHINI, G. Molecular mechanisms of human t-cell leukemia/lymphotropic virus type i infection. **Blood,** v. 86, n. 10, p. 3619-3639, 1995.

FUKUMOTO, R. et al. Inhibition of t-cell receptor signal transduction and viral expression by the linker for activation of t cells-interacting p12(i) protein of human t-cell leukemia/lymphoma virus type 1. **Journal of Virology,** v. 81, n. 17, p. 9088-9099, 2007.

FURUKAWA, Y. et al. Different cytokine production in tax-expressing cells between patients with human t cell lymphotropic virus type i (htlv-i)-associated myelopathy/tropical spastic paraparesis and asymptomatic HTLV-I carriers. **Journal of Infectious Diseases,** v. 187, n. 7, p. 1116-1125, 2003.

GAUDRAY, G. et al. The complementary strand of the human t-cell leukemia virus type 1 rna genome encodes a bzip transcription factor that down-regulates viral transcription. **Journal of Virology,** v. 76, n. 24, p. 12813-12822, 2002.

GESSAIN, A. et al. Antibodies to human t-lymphotropic virus type-i in patients with tropical spastic paraparesis. **Lancet,** v. 2, n. 8452, p. 407-410, 1985.

GESSAIN, A.; CASSAR, O. Epidemiological aspects and world distribution of htlv-1 infection. **Frontiers in Microbiology,** v. 3, n. p. 2012.

GESSAIN, A.; DE THE, G. Geographic and molecular epidemiology of primate t lymphotropic retroviruses: HTLV-I, HTLV-II, STLV-I, STLV-PP, AND PTLV-L. **Adv Virus Res**. v. 47, n. p. 377-426, 1996.

GHEZ, D. et al. Neuropilin-1 is involved in human t-cell lymphotropic virus type 1 entry. **Journal of Virology,** v. 80, n. 14, p. 6844-6854, 2006.

GONÇALVES, D. U. et al. dermatologic lesions in asymptomatic blood donors seropositive for human t cell lymphotropic virus type-1. **American Journal of Tropical Medicine and Hygiene,** v. 68, n. 5, p. 562-565, 2003.

GONCALVES, D. U. et al. HTLV-1-associated myelopathy/tropical spastic paraparesis (ham/tsp) inflammatory network. **Inflammation and Allergy Drug Targets,** v. 7, n. 2, p. 98-107, 2008.

GREEN, P.L.; CHEN, I. S. Y. **Fields Virology.** 4 ed. Philadelphia : Lippincott Willians e Wilkins, 2001, 1941-1969.

HALIN, M. et al. Human t-cell leukemia virus type 2 produces a spliced antisense transcript encoding a protein that lacks a classic bzip domain but still inhibits tax2-mediated transcription. **Blood,** v. 114, n. 12, p. 2427-2438, 2009.

HANCHARD, B. et al. **Human Retrovirology: Htlv.** New York: Raven Press, 1990.

HANON, E. et al. Fratricide among cd8(+) t lymphocytes naturally infected with human t cell lymphotropic virus type I. **Immunity,** v. 13, n. 5, p. 657-664, 2000.

HAYASHI, D. et al. Accumulation of human t-lymphotropic virus type i (htlv-i)-infected cells in the cerebrospinal fluid during the exacerbation of htlv-i-associated myelopathy. **Journal of Neurovirology,** v. 14, n. 5, p. 459-463, 2008.

HINUMA, Y. et al. Adult T-Cell Leukemia: antigen in an atl cell line and detection of antibodies to the antigen in human sera. **Proceedings of the National Academy Sciience,** v. 78, n. 10, p. 6476-6480, 1981.

HOSHINO, H. Cellular factors involved in HTLV-1 entry and pathogenicit. **Frontiers in Microbiology**. v. 3, n. p. 2012.

JACOBSON, S. Immunopathogenesis of human t cell lymphotropic virus type i-associated neurologic disease. **Journal of Infectious Diseases,** v. 186 Suppl 2, n. p. S187-192, 2002.

JACOBSON, S. et al. Circulating cd8+ cytotoxic t lymphocytes specific for htlv-i px in patients with htlv-i associated neurological disease. **Nature,** v. 348, n. 6298, p. 245-248, 1990.

JAIN, P. et al. Dc-Sign mediates cell-free infection and transmission of human t-cell lymphotropic virus type 1 by dendritic cells. **Journal of Virology,** v. 83, n. 21, p. 10908-10921, 2009.

JOHNSON, J. M.; HARROD, R.; FRANCHINI, G. Molecular biology and pathogenesis of the human t-cell leukaemia/lymphotropic virus type-1 (htlv-1). **International Journal Experimental Pathology,** v. 82, n. 3, p. 135-147, 2001.

JONES, K. S. et al. Molecular aspects of htlv-1 entry: functional domains of the htlv-1 surface subunit (su) and their relationships to the entry receptors. **Viruses,** v. 3, n. 12, p. 794-810, 2011.

JONES, K. S. et al. Cell-Free HTLV-1 infects dendritic cells leading to transmission and transformation of cd4(+) t cells. **Nature Medicine,** v. 14, n. 4, p. 429-436, 2008.

KALYANARAMAN, V. S. et al. New subtype of human t-cell leukemia virus (htlv-ii) associated with a t-cell variant of hairy cell leukemia. **Science,** v. 218, n. 4572, p. 571-573, 1982.

KANNAGI, M. et al. Predominant recognition of human t cell leukemia virus type i (HTLV-I) px gene products by human cd8+ cytotoxic t cells directed against htlv-i-infected cells. **International Immunology,** v. 3, n. 8, p. 761-767, 1991.

KAPLAN, J. E. et al. The risk of development of htlv-i-associated myelopathy/tropical spastic paraparesis among persons infected with HTLV-I. **Journal of Acquired Immune Deficience Syndrome,** v. 3, n. 11, p. 1096-1101, 1990.

KIM, S. J. et al. A conserved calcineurin-binding motif in human t lymphotropic virus type 1 p12i functions to modulate nuclear factor of activated t cell activation. **Journal of Biological Chemistry,** v. 278, n. 18, p. 15550-15557, 2003.

KIM, S. J. et al. Enhancement of Lfa-1-mediated t cell adhesion by human t lymphotropic virus type 1 p12i1. **Journal of Immunology,** v. 176, n. 9, p. 5463-5470, 2006.

KNIGHT, S. C. Dendritic cells in HIV-1 and HTLV-1 infection. **Advances in Experimental Medicine and Biology,** v. 329, n. p. 545-549, 1993.

KORALNIK, I. J. et al. Mapping of the intermolecular association of human t cell leukaemia/lymphotropic virus type i p12i and the vacuolar h+-atpase 16 kda subunit protein. **Journal of General Virology,** v. 76, Pt 8, n. p. 1909-1916, 1995.

KOYANAGI, Y. et al. In vivo infection of human t-cell leukemia virus type i in non-t cells. **Virology,** v. 196, n. 1, p. 25-33, 1993.

LAL, R. B. et al. Sensitivity and specificity of a recombinant transmembrane glycoprotein (rgp21)-spiked western immunoblot for serological confirmation of human t-cell

lymphotropic virus type i and type ii infections. **Journal of Clinical Microbiology,** v. 30, n. 2, p. 296-299, 1992.

LEVIN, M. C. et al. Autoimmunity due to molecular mimicry as a cause of neurological disease. **Nature Medicine,** v. 8, n. 5, p. 509-513, 2002.

MAHIEUX, R. A vaccine against HTLV-1 HBZ makes sense. **Blood,** v.126, n. 9, p. 1052-3, Ago. 2015.

MANEL, N. Human t cell leukemia virus envelope binding and virus entry are mediated by distinct domains of the glucose transporter Glut1. **Journal of Biological Chemistry,** v. 280, n. 32, p. 29025-29029, 2005.

MATSOUKA, M. Human t-cell leukemia virus type i (htlv-i) infection and the onset of adult t-cell leukemia (atl). **Retrovirology,** v. 2, n. 1, p. 27, 2005.

MATSOUKA, M. GREEN, P. L. The HBZ gene, a key player in HTLV-1 pathogenesis. **Retrovirology,** v. 6, n. 1, p. 71, 2009.

MATSUZAKI, T. et al. HTLV-I proviral load correlates with progression of motor disability in HAM/TSP: analysis of 239 HAM/TSP patients including 64 patients followed up for 10 years. **Journal of Neurovirology,** v.7, n. 3, p. 228–234, 2001.

MATSUOKA, M.; JEANG, K-T. Human t-cell leukaemia virus type 1 (htlv-1) infectivity and cellular transformation. **Nature Reviews Cancer,** v. 7, n. 4, p. 270-280, 2007.

MCCALLUM, R.M. et al. Arthritis syndromes associated with human t cell lymphotropic virus type i infection. **Medical Clinics of North American,** v. 81, n. 1, p. 261-276, 1997.

MESNARD, J. M.; BARBEAU, B. DEVAUX, C. HBZ, a new important player in the mystery of adult t-cell leukemia. **Blood,** v. 108, n. 13, p. 3979-3982, 2006.

MOCHIZUIKI, M. et al. Uveitis associated with human t lymphotropic virus type i: seroepidemiologic, clinical, and virologic studies. **Journal of Infectious Diseases,** v. 166, n. 4, p. 943-944, 1992.

MONTANHEIRO, P. A. The elevated interferon gamma production is an important immunological marker in ham/tsp pathogenesis. **Scandnavian Journal of Immunology,** v. 70, n. 4, p. 403-407, 2009.

MORGAN, O. S. et al. HTLV-1 and polymyositis in Jamaica. **Lancet,** v. 2, n. 8673, p. 1184-1187, 1989.

MOTA-MIRANDA, A. C. et al. Molecular characterization of htlv-1 gp46 glycoprotein from health carriers and ham/tsp infected individuals. **Virology Journal,** v. 10, n. p. 75, 2013a.

MOTA-MIRANDA, A. C. et al. Molecular study of hbz and gp21 human t cell leukemia virus type 1 proteins isolated from different clinical profile infected individuals. **AIDS Research Human Retroviruses,** v. 29, n. 10, p. 1370-1372, 2013b.

MUELLER, N. The epidemiology of HTLV-I infection. **Cancer Causes Control**, v. 2, n. 1, p. 37-52, 1991.

MUELLER, N. et al. Findings from the Miyazaki Cohort Study. **Journal of Acquired Immune Deficience Syndromes Human Retrovirology,** v. 13 Suppl 1, n. p. S2-7, 1996.

MUNIZ, A. L. et al. Association of cytokines, neurological disability, and disease duration in HAM/TSP patients. **Arquivos de Neuropsiquiatria,** v. 64, n. 2A, p. 217-221, 2006.

NAIR, A. M. et al. Calcium-dependent enhancement of transcription of p300 by human t-lymphotropic type 1 P12i. **Virology,** v. 353, n. 2, p. 247-257, 2006.

NAKAMURA, S. et al. Detection of tumor necrosis factor-alpha-positive cells in cerebrospinal fluid of patients with htlv-i-associated myelopathy. **Journal of Neuroimmunology,** v. 42, n. 2, p. 127-130, 1993.

NASCIMENTO, C. R. et al. Monocytes from HTLV-1-infected patients are unable to fully mature into dendritic cells. **Blood,** v. 117, n. 2, p. 489-499, 2011.

NATH, M. D. Regulation of the cell-surface expression of an htlv-i binding protein in human t cells during immune activation. **Blood,** v. 101, n. 8, p. 3085-3092, 2002.

NEJMEDDINE, M.; BANGHAM, C. R. M. The HTLV-1 Virological Synapse. **Viruses,** v. 2, n. 7, p. 1427-1447, 2010.

NISHIMURA, M. et al. Tumor necrosis factor, tumor necrosis factor receptors type 1 and 2, lymphotoxin-alpha, and hla-drb1 gene polymorphisms in human t-cell lymphotropic virus type i associated myelopathy. **Human Immunology,** v. 61, n. 12, p. 1262-1269, 2000.

NISHIOKA, K. HTLV-I Arthropathy and Sjogren Syndrome. **Journal of Acquired Immune Deficinces Syndrome Human Retrovirology,** v. 13 Suppl 1, n. p. S57-62, 1996.

NISHIOKA, K. et al. Chronic inflammatory arthropathy associated with HTLV-I. **Lancet,** v. 1, n. 8635, p. 441, 1989.

OSAME, M. Pathological mechanisms of human t-cell lymphotropic virus type i-associated myelopathy (Ham/Tsp). **Journal of Neurovirology,** v. 8, n. 5, p. 359-364, 2002.

OSAME, M. et al. HTLV-I associated myelopathy, a new clinical entity. **Lancet,** v. 1, n. 8488, p. 1031-1032, 1986.

PAIS-CORREIA, A-M. et al. Biofilm-like extracellular viral assemblies mediate htlv-1 cell-to-cell transmission at virological synapses. **Nature Medicine,** v. 16, n. 1, p. 83-89, 2009.

PINON, J. D. et al. Human t-cell leukemia virus type 1 envelope glycoprotein gp46 interacts with cell surface heparan sulfate proteoglycans. **Journal of Virology,** v. 77, n. 18, p. 9922-9930, 2003.

PISE-MASISON, C. A. et al. Co-Dependence of HTLV-1 p12 and p8 functions in virus persistence. **PLoS Pathogens,** v. 10, n. 11, p. e1004454, 2014.

POIESZ, B. J. et al. Detection and isolation of type c retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous t-cell lymphoma. **Proceedings National Academy of Science,** v. 77, n. 12, p. 7415-7419, 1980.

PRESTON, B. D.; DOUGHERTY, J. P. Mechanisms of retroviral mutation. **Trends Microbiology,** v. 4, n. 1, p. 16-21, 1996.

PROIETTI, A. B. F. C. **Cadernos Hemominas, 8.** 4 ed. Fundação Centro de Hematologia e Hemoterapia de Minas Gerais. Belo Horizonte, 2006.

RENDE, F. et al. Comparison of the genetic organization, expression strategies and oncogenic potential of HTLV-1 and HTLV-2. **Leukemia Research and Treatment,** v. 2012, p. 1-14, 2012.

RODRÍGUEZ, S. M. et al. Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV. **Viruses,** v. 3, n. 7, p. 1210-48, Jul. 2011.

ROMAN, G. C.; OSAME, M. Identity of HTLV-I-associated tropical spastic paraparesis and htlv-i-associated myelopathy. **Lancet,** v. 1, n. 8586, p. 651, 1988.

SANTOS, F. L. N.; LIMA, F. W. D. M. Epidemiologia, fisiopatogenia e diagnóstico laboratorial da infecção pelo HTLV-I. **Jornal Brasileiro de Patologia e Medicina Laboratorial,** v. 41, n. 2, p. 105-116, 2005.

SEIKI, M. et al. Human adult t-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell dna. **Proceedings National Academy of Science,** v. 80, n. 12, p. 3618-3622, 1983.

SETOYAMA, M. et al. Detection of HTLV-1 by polymerase chain reaction in situ hybridization in adult t-cell leukemia/lymphoma. **American Journal Pathology,** v. 152, n. 3, p. 683-689, 1998.

SHERMAN, M. P. et al. Sequence analysis of an immunogenic and neutralizing domain of the human t-cell lymphoma/leukemia virus type i gp46 surface membrane protein among various primate t-cell lymphoma/leukemia virus isolates including those from a patient with both htlv-i-associated myelopathy and adult t-cell leukemia. **Cancer Research,** v. 53, n. 24, p. 6067-6073, 1993.

SHIMOYAMA, M. Diagnostic criteria and classification of clinical subtypes of adult t-cell leukaemia-lymphoma. a report from the lymphoma study group (1984-87). **Bristish Journal Haematology,** v. 79, n. 3, p. 428-437, 1991.

SMITH, M. R.; GREENE, W. C. Characterization of a novel nuclear localization signal in the htlv-i tax transactivator protein. **Virology,** v. 187, n. 1, p. 316-320, 1992.

TSCHACHLER, E. et al. Human t-lymphotropic virus type i tax regulates the expression of the human lymphotoxin gene. **Blood,** v. 81, n. 1, p. 95-100, 1993.

VAN PROOYEN, N. et al. Human t-cell leukemia virus type 1 p8 protein increases cellular conduits and virus transmission. **Proceedings National Academy of Science,** v. 107, n. 48, p. 20738-20743, 2010.

VERDONCK, K. et al. Human t-lymphotropic virus 1: recent knowledge about an ancient infection. **The Lancet Infectious Diseases,** v. 7, n. 4, p. 266-281, 2007.

WATTEL, E. et al. Clonal expansion of human t-cell leukemia virus type i-infected cells in asymptomatic and symptomatic carriers without malignancy. **J. Virology,** v. 69, n. 5, p. 2863-2868, 1995.

WOLFE, N. D. et al. Emergence of unique primate t-lymphotropic viruses among central african bushmeat hunters. **Proceedings of the National Academy of Sciences,** v. 102, n. 22, p. 7994-7999, 2005.

XIE, L.; GREEN, P. L. Envelope is a major viral determinant of the distinct in vitro cellular transformation tropism of human t-cell leukemia virus type 1 (Htlv-1) and Htlv-2. **Journal of Virology,** v. 79, n. 23, p. 14536-14545, 2005.

YAMAGUCHI, K. Human T-Lymphotropic Virus Type I in Japan. **Lancet,** v. 343, n. 8891, p. 213-216, 1994.

YAMANO, Y. et al. Correlation of human t-cell lymphotropic virus type 1 (htlv-1) mrna with proviral dna load, virus-specific cd8(+) t cells, and disease severity in htlv-1-associated myelopathy (Ham/Tsp). **Blood,** v. 99, n. 1, p. 88-94, 2002.

YASUNAGA, J.; MATSOUKA, M. Human t-cell leukemia virus type i induces adult t-cell leukemia: from clinical aspects to molecular mechanisms. **Cancer Control,** v. 14, n. 2, p. 133-140, 2007.

YOSHIDA, M.; MIYOSHI, I.; HINUMA, Y. Isolation and characterization of retrovirus from cell lines of human adult t-cell leukemia and its implication in the disease. **Proceedings of the National Academy of Sciences,** v. 79, n. 6, p. 2031-2035, 1982.

YOUNIS, I.; GREEN, P. L. The Human T-Cell Leukemia Virus Rex Protein. **Frontiers in Bioscience,** v. 10, n. p. 431-445, 2005.

APÊNDICE

Durante o desenvolvimento desta tese alguns projetos foram realizados em paralelo e encontram-se publicados em revistas indexadas ou em processo de submissão.

Mem Inst Oswaldo Cruz, Rio de Janeiro: 1-4, 2014 1

Inferences about the global scenario of human T-cell lymphotropic virus type 1 infection using data mining of viral sequences

Thessika Hialla Almeida Araujo[1], Fernanda Khouri Barreto[1], Luiz Carlos Júnior Alcântara[1]/+, Aline Cristina Andrade Mota Miranda[2]

[1]Centro de Pesquisa Gonçalo Moniz-Fiocruz, Salvador, BA, Brasil
[2]Instituto de Ciências da Saúde, Universidade Federal da Bahia, Salvador, BA, Brasil

Human T-cell lymphotropic virus type 1 (HTLV-1) is mainly associated with two diseases: tropical spastic paraparesis/HTLV-1-associated myelopathy (TSP/HAM) and adult T-cell leukaemia/lymphoma. This retrovirus infects five-10 million individuals throughout the world. Previously, we developed a database that annotates sequence data from GenBank and the present study aimed to describe the clinical, molecular and epidemiological scenarios of HTLV-1 infection through the stored sequences in this database. A total of 2,545 registered complete and partial sequences of HTLV-1 were collected and 1,967 (77.3%) of those sequences represented unique isolates. Among these isolates, 93% contained geographic origin information and only 39% were related to any clinical status. A total of 1,091 sequences contained information about the geographic origin and viral subtype and 93% of these sequences were identified as subtype "a". Ethnicity data are very scarce. Regarding clinical status data, 29% of the sequences were generated from TSP/HAM and 67.8% from healthy carrier individuals. Although the data mining enabled some inferences about specific aspects of HTLV-1 infection to be made, due to the relative scarcity of data of available sequences, it was not possible to delineate a global scenario of HTLV-1 infection.

Key words: HTLV-1 - data mining - HTLV-1 database

Human T-cell lymphotropic virus type 1 (HTLV-1) was the first described human retrovirus (Poiesz et al. 1980). This retrovirus is the causative agent of tropical spastic paraparesis/HTLV-1-associated myelopathy (TSP/HAM) (Gessain et al. 1985), adult T-cell leukaemia/lymphoma (ATL) (Yoshida et al. 1982) and other inflammatory diseases such as HTLV-1-associated infectious dermatitis (La Grenade et al. 1998) and HTLV-1-associated uveitis (Mochizuki et al. 1992). However, the pathogenesis of some clinical manifestations is not yet fully understood.

Epidemiological data show that HTLV-1 has a worldwide distribution and it is estimated that five-10 million people are infected (Gessain & Cassar 2012). This infection is endemic in southwestern Japan (Mueller et al. 1996), sub-Saharan Africa (Gessain & de Thé 1996), regions of the Caribbean (Hanchard et al. 1990) and minor areas in Iran, Melanesia (Mueller 1991) and Brazil (Galvão-Castro et al. 1997).

Regardless, HTLV-1 epidemiology still presents many challenges. Virus prevalence rates have been correlated with geographic characteristics and the social setting of destitute populations. However, these populations are not frequently the target of great public and government interest (Galvão-Castro et al. 1997). Molecular studies, especially during the late decade, have contributed to the acquisition of knowledge about virus epidemiology and the molecular characteristics. Furthermore, a great amount of viral sequences are generated from these molecular studies because of this appropriate data management and data mining can provide additional consistent information about HTLV-1 infection.

In response to the need of obtaining more information about the already generated and available HTLV-1 sequences, HTLV-1 Molecular Epidemiology Database (htlv1db.fiocruz.bahia.br) was developed (Araujo et al. 2012). This database contains information that can support our understanding of viral pathogenesis, the route of transmission, polymorphisms, epidemiology, genotype-phenotype relationships, geographic distribution and viral evolution. Therefore, the purpose of the present study was to assess the different types of information deposited in HTLV-1 Molecular Epidemiology Database to describe clinical, molecular and epidemiological scenarios about HTLV-1 infection.

MATERIALS AND METHODS

This is a descriptive study about the clinical, molecular and epidemiological data of HTLV-1 infection that are associated with the stored genetic sequences in HTLV-1 Molecular Epidemiology Database.

All the descriptive analyses were performed using the search algorithm implemented at the HTLV-1 database (Araujo et al. 2012). Initially, we made a list with the variables (age, gender, clinical status, subtype, subgroup, geographic origin) that were more frequent in the database. We then performed combinations with the listed variables; for example, we searched for sequences with information about geographic origin, viral subtype and clinical status. These combinations constitute the

doi: 10.1590/0074-0276130587
+ Corresponding author: alcan@bahia.fiocruz.br
Received 18 December 2013
Accepted 7 March 2014

Molecular characterization of new HTLV-1 complete genomes from patients with different clinical forms

Thessika Hialla Almeida Araújo[1]; Aline Dórea Luz Menezes1; Fernanda Khouri Barreto[1]; Clayton Pereira Silva de Lima[4]; Rodrigo Santos de Oliveira[4]; Poliana da Silva Lemos[4]; Bernardo Galvão-Castro[5]; Simone Kashima Haddad[6]; Maria Lourdes Farre Vallve[7]; Achilea Candida Lisboa Bittencourt [9]; Edgar Marcelino de Carvalho Filho[1,9] ; Luciane Amorim Santos[2]; Filipe Ferreira de Almeida Rego[3]; Adjile Edjide Roukiyath Amoussa[1]; Marta Giovanetti[1,10]; Aline Cristina Andrade Mota Miranda Mascarenhas[8]; Márcio Roberto Teixeira Nunes[4]; Luiz Carlos Júnior Alcântara[1,*].

Abstract

The HTLV-1 is the first human retrovirus and is associated with several clinical syndromes, however, the pathogenesis of these clinical manifestations is still not fully understood. Furthermore, there are few complete genomes publicly available, about 0.12 complete genomes per 10,000 infected individuals and the databases have a major deficiency of sequences information. This study generate and characterized 31 HTLV-1 complete genomes sequences derived from individuals with TSP/HAM, ATL, HID and asymptomatic patients, whose sequences have the information clinical and epidemiological about the patients. The sequencing data were assembled and mapped against the reference HTLV-1 genome. These sequences were genotyped as Cosmopolitan subtype, Transcontinental subgroup. We identified the variants in the coding regions of the genome of the different clinical profiles, however, no statistical relation was detected.In a second phase of the study, we searched the mutations found in the sequences generated in sequences previously published in Genbank, corroborating previous data, there was no statistical significance. This study contributed to increase of HTLV-1 complete genomes in world. Furthermore, to investigate better the contribution of HTLV-1 mutations for the disease outcome it is necessary evaluate the interaction of the viral genome and characteristics of the human host.

Keywords: HTLV-1; complete genome; next generation sequencing.

Differential gene expression in HTLV-1 infected patients with HAM/TSP: A meta-analysis of transcriptomic data

Marcia Weber Carneiro[a]; Maria Fernanda de Castro-Amarante[a]; Fernanda Khouri Barreto[a]; Vagner de Souza Fonseca[a]; Paloma Viana [a]; Luiz Carlos Junior Alcantara[a,*]

Abstract

Most patients infected with the human T-cell lymphotropic virus type 1 efficiently control the infection and remains clinically asymptomatic while about 1-5% may develop an aggressive malignancy of $CD4^+$ T lymphocytes, known as adult T-cell leukemia/lymphoma (ATLL), or a tropical spastic paraparesis/HTLV-associated myelopathy (HAM/TSP). The molecular mechanisms used by the virus to establish persistent infection and subsequently promote lymphocyte proliferation while evading elimination by the host immune response remain poorly defined. To better understand the potential biological process involved in the HAM/TSP development we performed bioinformatics analysis using gene expression data from the Gene Expression Omnibus (GEO) database. Herein, we compared 5 GEO datasets from patients with HAM/TSP, asymptomatic HTLV-1 carrier (AHC) and healthy controls (HC) in $CD4^+$ T cells or whole blood. We found 13 genes in common in the platforms differently expressed in the different groups. These genes were associated with upstream regulators genes, infectious diseases, inflammatory response, cancer, hereditary disorder, neurological disease, and important molecular and cellular functions. This study help understanding the underlying molecular pathways related to pathogenesis of this infection in association with those common target genes, that may serve as potential biomarkers.

Keywords: HTLV-1; Gene expression; HAM/TSP; GEO.